# STIC Search Report
## Biotech-Chem Library

USPTO

TO: Manjunath N Rao
Location: CM!/10A11/10D01
Art Unit: 1652
Thursday, May 29, 2003
Case Serial Number: 922683

From: Paul Schulwitz
Location: Biotech-Chem Library
CM1-6B06
Phone: 305-1954

paul.schulwitz@uspto.gov

## Search Notes

Examiner Rao,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz
Technical Information Specialist
STIC Biotech/Chem Library
(703)305-1954

## STIC-Biotech/ChemLib

**From:** Rao, Manjunath N.
**Sent:** Wednesday, May 21, 2003 1:54 PM
**To:** STIC-Biotech/ChemLib
**Subject:** Sequence search request for 09/922683

```
From: Manjunath N. Rao
      Art Unit 1652, Room 10A11
      Mail Box in Room 10D 01
      Phone: 306-5681
```

**Date: 5-21-03**

Please search the following as soon as possible for application with serial number **09/922683**

**1. SEQ ID NO: 7,** against all <u>commercial nucleic acid databases</u> including <u>issued patents database</u> and <u>pending application database</u> and provide **a print of all results.**

**If you have any questions please call me at the above phone number.**

Thanks

**Manjunath N. Rao, Ph.D.**
**Biotechnology Patent Examiner**
**Art Unit 1652, Room 10A11**
**Mail Box in 10D01**
**Crystal Mall 1, USPTO.**

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

| | TYPE OF SEARCH: | VENDOR/COST (where applic.) |
|---|---|---|
| Searcher:_____ | NA Sequences:_____ | STN:_____ |
| Phone:_____ | AA Sequences:_____ | DIALOG:_____ |
| Location:_____ | Structures:_____ | Questel/Orbit:_____ |
| Date Picked Up:___5/23 | Bibliographic:_____ | DRLink:_____ |
| Date Completed:__5/29 | Litigation:_____ | Lexis/Nexis:_____ |
| Searcher Prep/Review:_____ | Full text:_____ | Sequence Sys.:_____ |
| Clerical:_____ | Patent Family:_____ | WWW/Internet:_____ |
| Online time:_____ | Other:_____ | Other (specify):_____ |

OM nucleic - nucleic search, using sw model

Run on:
   GenCore version 5.1.4_p5_4578
   Copyright (c) 1993 - 2003 Compugen Ltd

Run on: May 28, 2003, 15:28:40 ; Search time 11321 Seconds
           (without alignments)
         17619.524 Million cell updates/sec

Title:   US-09-922-683-7
Perfect score: 6854
Sequence: 1 CTGCAGGGTCCCTGGTGCA........TCGGAGCCGATCCGGCTGCAG 6854

Scoring table: IDENTITY_NUC
      Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
      Maximum Match 100%
      Listing first 45 summaries

Database : GenEmbl:*
  1: gb_ba:*
  2: gb_htg:*
  3: gb_in:*
  4: gb_om:*
  5: gb_ov:*
  6: gb_pat:*
  7: gb_ph:*
  8: gb_pl:*
  9: gb_pr:*
 10: gb_ro:*
 11: gb_sts:*
 12: gb_sy:*
 13: gb_un:*
 14: gb_vi:*
 15: em_ba:*
 16: em_fun:*
 17: em_hum:*
 18: em_in:*
 19: em_mu:*
 20: em_om:*
 21: em_or:*
 22: em_ov:*
 23: em_pat:*
 24: em_ph:*
 25: em_pl:*
 26: em_ro:*
 27: em_sts:*
 28: em_un:*
 29: em_vi:*
 30: em_htg_hum:*
 31: em_htg_inv:*
 32: em_htg_other:*
 33: em_htg_mus:*
 34: em_htg_pln:*
 35: em_htg_rod:*
 36: em_htg_mam:*
 37: em_htg_vrt:*
 38: em_sy:*
 39: em_htgo_hum:*
 40: em_htgo_mus:*
 41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

## SUMMARIES

| | Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| | 1 | 6854 | 100.0 | 6854 | 6 | A93103 | A93103 Sequence 7 |
| | 2 | 6854 | 100.0 | 6854 | 6 | AR174057 | AR174057 Sequence |
| | 3 | 986.2 | 14.4 | 9523 | 1 | AF237894S1 | AF237894 Streptomy |
| | 4 | 977.8 | 14.3 | 2404 | 1 | SAMTMDE | Y10907 S.argillace |
| | 5 | 977.8 | 14.3 | 18977 | 1 | SAR7932 | AJ007932 Streptomy |
| | 6 | 973.4 | 14.2 | 15052 | 1 | AF055579 | AF055579 Streptomy |
| | 7 | 903.2 | 13.2 | 25459 | 1 | SGAJ6985 | AJ006985 Streptomy |
| c | 8 | 877.8 | 12.8 | 109519 | 6 | AX195929 | AX195929 Sequence |
| | 9 | 875.4 | 12.8 | 2682 | 1 | STMDNRLM | L47163 Streptomyce |
| | 10 | 867.2 | 12.7 | 34644 | 1 | AF080235 | AF080235 Streptomy |
| c | 11 | 855 | 12.5 | 5010 | 1 | SCMALREFG | Y07706 S.coelicolo |
| | 12 | 855 | 12.5 | 33517 | 1 | SC10B7 | AL355752 Streptomy |
| | 13 | 850.8 | 12.4 | 39250 | 1 | SVI011500 | AJ011500 Streptomy |
| | 14 | 845.4 | 12.3 | 8018 | 1 | AF164960 | AF164960 Streptomy |
| c | 15 | 834.6 | 12.2 | 7365 | 1 | STMSVIDEOX | L37334 Streptomyce |
| | 16 | 817.2 | 11.9 | 14186 | 6 | AX204987 | AX204987 Sequence |
| | 17 | 763.8 | 11.1 | 59816 | 1 | AF333038 | AF333038 Streptomy |
| | 18 | 746.6 | 10.9 | 1800 | 1 | AF144042 | AF144042 Streptomy |
| c | 19 | 621 | 9.1 | 1460 | 1 | SLI272397 | AJ272397 Streptomy |
| c | 20 | 612.4 | 8.9 | 1435 | 1 | SLREG1 | X98242 S.lividans |
| c | 21 | 575.2 | 8.4 | 34074 | 1 | SCF81 | AL133171 Streptomy |
| | 22 | 542.6 | 7.9 | 4062 | 1 | AF128273 | AF128273 Streptomy |
| | 23 | 536.2 | 7.8 | 90348 | 1 | AF497482 | AF497482 Micromono |
| c | 24 | 527.4 | 7.7 | 12441 | 1 | AF079762 | AF079762 Streptomy |
| c | 25 | 523.8 | 7.6 | 2242 | 1 | SLKAGLA | Y08304 S.limosus a |
| | 26 | 522.6 | 7.6 | 546 | 6 | A93099 | A93099 Sequence 3 |
| | 27 | 522.6 | 7.6 | 546 | 6 | AR174055 | AR174055 Sequence |
| | 28 | 503.2 | 7.3 | 4094 | 1 | AF126354 | AF126354 Streptomy |
| c | 29 | 497.6 | 7.3 | 2634 | 6 | A25110 | A25110 S. nodosus |
| c | 30 | 497.6 | 7.3 | 2634 | 6 | I38748 | I38748 Sequence 31 |
| | 31 | 497.6 | 7.3 | 2634 | 6 | I81357 | I81357 Sequence 31 |
| | 32 | 486.8 | 7.1 | 16065 | 1 | AF187532 | AF187532 Streptomy |
| c | 33 | 486.4 | 7.1 | 5084 | 1 | SGSTRELMB | X62567 S.griseus s |
| | 34 | 485.6 | 7.1 | 1065 | 6 | AX196000 | AX196000 Sequence |
| | 35 | 485.2 | 7.1 | 17665 | 1 | AF521878 | AF521878 Streptomy |
| c | 36 | 483.4 | 7.1 | 12203 | 1 | SGSTR | Y00459 Streptomyce |
| | 37 | 475 | 6.9 | 2689 | 1 | AF128272 | AF128272 Streptomy |
| | 38 | 473.4 | 6.9 | 10167 | 1 | SFU08223 | U08223 Streptomyce |
| | 39 | 469 | 6.8 | 1132 | 1 | AF306787 | AF306787 Streptomy |
| | 40 | 465.8 | 6.8 | 990 | 6 | AX089467 | AX089467 Sequence |
| | 41 | 465.8 | 6.8 | 2310 | 1 | AR165019 | AR165019 Sequence |
| | 42 | 465.8 | 6.8 | 2311 | 1 | AF355468 | AF355468 Saccharop |
| | 43 | 465.8 | 6.8 | 2363 | 6 | AX089466 | AX089466 Sequence |
| | 44 | 454.4 | 6.6 | 3035 | 1 | AF201913 | AF201913 Streptomy |
| c | 45 | 454.2 | 6.6 | 1476 | 6 | AR173233 | AR173233 Sequence |

## ALIGNMENTS

RESULT 1
A93103

```
LOCUS       A93103                  6854 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9747748.
ACCESSION   A93103
VERSION     A93103.1  GI:6741499
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified.
            unclassified.
REFERENCE   1  (bases 1 to 6854)
  AUTHORS   Decker,H.
  TITLE     ISOLATION OF THE BIOSYNTHESIS GENES FOR PSEUDO-OLIGOSACCHARIDES
            FROM STREPTOMYCES GLAUCESCENS GLA.O AND THEIR USE
  JOURNAL   Patent: WO 9747748-A 7 18-DEC-1997;
```

Pred. No. is the number of results predicted by chance to have a

```
FEATURES       HOECHST AG (DE); DECKER HEINRICH (DE)
source         Location/Qualifiers
               1..6854
                 /organism="unidentified"
                 /db_xref="taxon:32644"
BASE COUNT     1018 a    2445 c    2445 g     946 t
ORIGIN

Query Match    100.0%;  Score 6854;  DB 6;  Length 6854;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 6854;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0:
```

```
Qy      1 CTGCAGGGTTCCCTGGTCGACGACCCGCCCCTGTCGACGACCAGGCGGCTGTCGCGACGAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTGCAGGGTTCCCTGGTCGACGACCCGCCCCTGTCGACGACCAGGCGGCTGTCGCGACGAT 60

Qy     61 GGGGCGATGTCGGCGATCGTCGTGCTGGCTGAGCACCACGGTGCCCAGTTCCCGTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGGGCGATGTCGGCGATCGTCGTGCTGGCTGAGCACCACGGTGCCCAGTTCCCGTG 120

Qy    121 GGGCGGGTTGACCAGCCGGCGCACCGCGCACCGCGTCCTTCAGCACCATGTCGAGGCCGATCGTGGG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGCGGGTTGACCAGCCGGCGCACCGCGCACCGCGTCCTTCAGCACCATGTCGAGGCCGATCGTGGG 180

Qy    181 CTCGTCCCAGAACACAGCACGGCCCGGTCGTGCAGCAGGCTCGCCCGCATCTCGGCGCCGCCAT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTCGTCCCAGAACACAGCACGGCCCGGTCGTGCAGCAGGCTCGCCCGCATCTCGGCGCCGCCAT 240

Qy    241 GGCTGTCCGAGGCTGAGCTGCCCACGGGGTGGACCCCAGCCGGCGTCGATGTCGAGGAG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCTGTCCGAGGCTGAGCTGCCCACGGGGTGGACCCCAGCCGGCGTCGATGTCGAGGAG 300

Qy    301 GTCCCGGACACAGGGCGAGGTTGCCCGGTAGACCGGTCGGGGATGTCGTAGATGCGGGCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTCCCGGACACAGGGCGAGGTTGCCCGGTAGACCGGTCGGGGATGTCGTAGATGCGGGCG 360

Qy    361 CAGGATGCCGAAGGAGTCGGGTACCGACAGGTCCACCAGAGCTGGCTGCGCGTGGCCGAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGGATGCCGAAGGAGTCGGGTACCGACAGGTCCACCAGAGCTGGCTGCGCGTGGCCGAA 420

Qy    421 GACGACGCCGATCGTGCGGGCGGTTGCGCTGCCGGTGCGCGGTAGGCGCGGTGGCCGGT 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GACGACGCCGATCGTGCGGGCGGTTGCGCTGCCGGTGCGCGGTAGGCGCGGTGGCCGGT 480

Qy    481 CGTGCAGCGGCGGCCGGAGGTGGGGTCATGATGCCGGTCAGCATCTTGATCGTGGTCGACTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CGTGCAGCGGCGGCCGGAGGTGGGGTCATGATGCCGGTCAGCATCTTGATCGTGGTCGACTT 540

Qy    541 CGTGCAGCGGCCGGAGGTGGGGTCTTGGGGTCAGCATCTTGATCGTGGTCGACTT 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCCGGCCTCGTTGCCGATGTGGGGTCTTGCCGGTGTACCGGTCAGCATCTTGATCGTGGTCGACTT 600

Qy    601 GTCGTCGACGCGGCCGCCACGACGGTCAGGAGGTGGAGGGTGCCGCGTGCCGCAG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTCGTCGACGCGGCCGCCACGACGGTCAGGAGGTGGAGGGTGCCGCGTGCCGCAG 660

Qy    661 CAGGATCACCCGCTCGCGTTCGACGGCGCCTCGTCGACGGCGGCTCG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CAGGATCACCCGCTCGCGTTCGACGGCGCCTCGTCGACGGCGGCTCG 720

Qy    721 GCGATCACCCGCTCGCGTTCGACGGCGCCTCGTCGACGGCGGCTCG 780
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCGATCACCCGCTCGCGTTCGACGGCGCCTCGTCGACGGCGGCTCG 780

Qy    781 TCGAGGGTGAACGGCGGGTGCCAGCGGCAGGATGTGCCGCCAGGGGCGCCAGCCCC 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TCGAGGGTGAACGGCGGGTGCCAGCGGCAGGATGTGCCGCCAGGGGTGCGCCAGCCCC 840

Qy    841 AGGTCGAGGCGGGTGTGTAGACGGCGGGTCTCGGGGCGTCCCGGCCGACG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGTCGAGGCGGGTGTGTAGACGGCGGGTCTCGGGGCGTCCCGGCCGACG 900

Qy    901 GCGTCGGTGACGAACTCCAGGCCCCACAGCAGTCCGAGCCCGCGCTACCTGGCCGAGCTGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GCGTCGGTGACGAACTCCAGGCCCCACAGCAGTCCGAGCCCGCGCTACCTGGCCGAGCTGG 960

Qy    961 GGGAAGCGGGACTCCAGGCGGCGCAGCCGCGCCGCCCTCGAGCTGGATGAGCTCGCCGAGGACGCGCACG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GGGAAGCGGGACTCCAGGCGGCGCAGCCGCGCCGCCCTCGAGCTGGATGAGCTCGCCGAGGACGCGCACG 1020

Qy   1021 CGGTCGATCAGCCGGTCGCGCTCGACGACTCGAGCCTGGCGGCGGGCGGCGGCGATCCCC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGGTCGATCAGCCGGTCGCGCTCGACGACTCGAGCCTGGCGGCGGGCGGCGGCGATCCCC 1080

Qy   1081 AGTGGGTTGCTCGCGTACTCGAGGCGTACGCCCGGGTGGCCGCCTCGGCGCTCCGGCTGCCA 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AGTGGGTTGCTCGCGTACTCGAGGCGTACGCCCGGGTGGCCGCCTCGGCGCTCCGGCTGCCA 1140

Qy   1141 GCTTCCGGCGTCCGGCCAGCAGCGGGAATCCGCTCCGGCGTGCCCTTGGACAGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GCTTCCGGCGTCCGGCCAGCAGCGGGAATCCGCTCCGGCGTGCCCTTGGACAGC 1200

Qy   1201 ATCGGCCAGTCCGGCTCGATGCCGAACAGTTCGCTGGCCGATCGTGGGGTGCGCCCG 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATCGGCCAGTCCGGCTCGATGCCGAACAGTTCGCTGGCCGATCGTGGGGTGCGCCCG 1260

Qy   1261 CCGCCGGTGAGGACCTCGTCGGCGACGAGCAGCACGCCCGTCCGGCAGGCGCGGCG 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCGCCGGTGAGGACCTCGTCGGCGACGAGCAGCACGCCCGTCCGGCAGGCGCGGCG 1320

Qy   1321 ATCCGCTCCCAGTAGCCGGTGCCGCTGAAGGCCGTGTCGGTGGCGCTGCCGAGGAG 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ATCCGCTCCCAGTAGCCGGTGCCGCTGAAGGCCGTGTCGGTGGCGCTGCCGAGGAG 1380

Qy   1381 AAGACCAGGCCGGAGACGTTGGGCTTCTCCGGGATGTGCCGGATGTCGTAGATGCGGGCG 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAGACCAGGCCGGAGACGTTGGGCTTCTCCGGGATGTGCCGGATGTCGTAGATGCGGGCG 1440

Qy   1441 CGCACGTCGCACGAGGGTACTCCAGGGGACAGCGGTAGCCAGTGGACCAGTAGGGCTGTA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CGCACGTCGCACGAGGGTACTCCAGGGGACAGCGGTAGCCAGTGGACCAGTAGGGCTGTA 1500

Qy   1501 GCCAGCACCGTGTTGCCGCTGAAGGCCTGTGTGGCCGATGTCCCAGTGGACCATCCGG 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCAGCACCGTGTTGCCGCTGAAGGCCTGTGTGGCCGATGTCCCAGTGGACCATCCGG 1560

Qy   1561 GCGGCCCATGTCTTGCCGTGAAGCCGGCCGTGGAGCCGGCCGCCTTGCGGCCCGGC 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GCGGCCCATGTCTTGCCGTGAAGCCGGCCGTGGAGCCGGCCGCAGATCGGCCCGGC 1620

Qy   1621 GCGGCCGTCGCCCTGGACGACCCAGAGGCGCGCCTCGACCACCTGGCCGACCAGTTCAGCAGGCCGACTT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCGGCCGTCGCCCTGGACGACCCAGAGGCGCGCCTCGACCACCTGGCCGACCAGTTCCAGCAGGCCGACTT 1680

Qy   1681 AAGGGCTAGGTGTCGAGCTCGTTCGGGCGGTGCTGCTCGTGTCCAGTTCCAGGGCCGGAAG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AAGGGCTAGGTGTCGAGCTCGTTCGGGCGGTGCTGCTCGTGTCCAGTTCCAGGGCCGGAAG 1740

Qy   1741 CGGTACGACCTTCCGGGTGGTCGAGCTCGTCGTTCCACAGCGGGTGGTTCCGGAGTGGTCGTGAGTGCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CGGTACGACCTTCCGGGTGGTCGAGCTCGTCGTTCCACAGCGGGTGGTTCCGGAGTGGTCGTGAGTGCC 1800

Qy   1801 TCGACGACTTCCTCCGGGTTGCGTCCGGGTGCCGGTGCCCCAGTGACTGGGCTGGGGTCCGGCGGCG 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TCGACGACTTCCTCCGGGTTGCGTCCGGGTGCCGGTGCCCCAGTGACTGGGCTGGGGTCCGGCGGCG 1860

Qy   1861 AGGTACTGGTTGCGCTGCGACCCTCCAGGTCGTCAGAACGGAACCGGCGTCCTCCGGG 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGGTACTGGTTGCGCTGCGACCCTCCAGGTCGTCAGAACGGAACCGGCGTCCTCCGGG 1920

Qy   1921 CGTCCCGGGACGCGGCTTCCTCCGGAGCCCGTCCCCGGCCTGGTGGCGGCCTCCCGTGCCGCAGG 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CGTCCCGGGACGCGGCTTCCTCCGGAGCCCGTCCCCGGCCTGGTGGCGGCCTCCCGTGCCGCAGG 1980

Qy   1981 TGCTGTGTCGTGCGGCTCATCCGGTAAGCCTGTCATCGCTGTCTCGTCGGACCGGCTGACGCGAT 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Left column:

```
Db  1981  TGCTGTGTCTGCCGTAAGCCTGTCATCGCTGTGCCTCTGCTCGTCGCTGCTCGTGACCGGCTGACGCGAT  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2041  CGCCGGCGAACTGCGTTGTGCGCACCACGGTTGGGCGGCTTCGGCGGCTGAGTCAAACAC  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  CGGCGGCGAACTGCGTTGTGCGCACCACGGTTGGGCGGCTTCGGCGGCTGAGTCAAACAC  2100

Qy  2101  TTGAACACACCGGCTGCAAGAGTTTGCGGGTTGTTTCAGAAAGTTGTTGCGAGCGGCCC  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  TTGAACACACCGGCTGCAAGAGTTTGCGGGTTGTTTCAGAAAGTTGTTGCGAGCGGCCCC  2160

Qy  2161  CGGCACTCTGGTTGAGTGACGTGTTACGGCGCACCACGCCTCACGTTCGAGGAGGGA  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  CGGCACTCTGGTTGAGTGACGTGTTACGGCGCACCACGCCTCACGTTCGAGGAGGGA  2220

Qy  2221  CCTGTGAGAACAAGCCCGCCAGACCCGCTCCCCGGAGGCGAGGTGAAGGCCCTGG  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  CCTGTGAGAACAAGCCCGCCAGACCCGCTCCCCGGAGGCGAGGTGAAGGCCCTGG  2280

Qy  2281  TCCTGGCAGGTGGAACCGGCAGCGGCAGACTGAGGCCGTTCACCCACCGCCGCCAAGCAGC  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  TCCTGGCAGGTGGAACCGGCAGCGGCAGACTGAGGCCGTTCACCCACCGCCGCCAAGCAGC  2340

Qy  2341  TGTCCCCATCGCCAACAAGCCCGTTCGGCTGCTCTTCTACGCCGTGAGTCCTCCGCCGGCGG  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TGCTCCCCATCGCCAACAAGCCCGTTCGGCTGCTCTTCTACGCCGTGAGTCCTCCGCCGGCGG  2400

Qy  2401  GTGTCCGGGAGGCCGGCGTCTGCGTGTCGTCGTCGGGAGATCCGGCGTACGGCGTTCCACCG  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  GTGTCCGGGAGGCCGGCGTCTGCGTGTCGTCGTCGGGAGATCCGGCGTACGGCGTTCCACCG  2460

Qy  2461  GCGACGGCACCGCGTTCGGGTTACGCATCACCTCACCTCACCGCCGCCCGCCCGCCGGTC  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GCGACGGCACCGCGTTCGGGTTACGCATCACCTCACCTCACCGCCGCCCGCCCGCCGGTC  2520

Qy  2521  TCGCCCACCACGGCGGTGCGCATCGCCCCGCGCGGCGTTCCTGGCGGACGACTTCCTGTGTACC  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  TCGCCCACCACGGCGGTGCGCATCGCCCCGCGCGGCGTTCCTGGCGGACGACTTCCTGTGTACC  2580

Qy  2581  TGGGGGACAACTACCTCGCCCAGGGCGTTCGGCGGTCACCGGCGTCCGCCAATCGGCCGCCGATC  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TGGGGGACAACTACCTCGCCCAGGGCGTTCGGCGGTCACCGGCGTCCGCCAATCGGCCGCCGATC  2640

Qy  2641  CCGCGGCGGCCCGGCGCTGCTCACCCCGGTCGCGCGACCCGTCGCCCTTCGGCGTCGGCGTCGCGG  2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  CCGCGGCGGCCCGGCGCTGCTCACCCCGGTCGCGCGACCCGTCGCCCTTCGGCGTCGGCGTCGCGG  2700

Qy  2701  AGGTCGACGCGGCGAACGGGCCTGCGCGTTGGAGGAGAAACCCGACGTCCCGCAGCT  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  AGGTCGACGCGGCGAACGGGCCTGCGCGTTGGAGGAGAAACCCGACGTCCCGCAGCT  2760

Qy  2761  CGCTCCGCGTCATCGGCGGTGTACGGCGTCACGTCTGGACGACGTCAGAGGCGGCATCGAGGGGA  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  CGCTCCGCGTCATCGGCGGTGTACGGCGTCACGTCTGGACGACGTCAGAGGCGGCATCGAGGGGA  2820

Qy  2821  TCACCCCTCGCCGCCGGAGCTGGAGATCACCGACGTGGAGTGATGATGACC  2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  TCACCCCTCGCCGCCGGAGCTGGAGATCACCGACGTGGAGTGATGATGACC  2880

Qy  2881  GGGGCCTGCGCGTACGGGGCCGAGACCAGCACCGCCTGCGCCTGGACACCGCG  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  GGGGCCTGCGCGTACGGGGCCGAGACCAGCACCGCCTGCGCCTGGACACCGCG  2940

Qy  2941  AGGACATGCTGGAGGTCAACCGGTCAACGTCACGTCGGCGGCGACTGAGGGGA  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  AGGACATGCTGGAGGTCAACCGGTCAACGTCACGTCGGCGGCGACTGAGGGGA  3000

Qy  3001  AGGTCGACGCGCAGCAGCACCACGTCTGGCGCGGACCGGCGTCGGCGGCCGCATCGTGC  3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  AGGTCGACGCGCAGCAGCACCACGTCTGGCGCGGACCGGCGTCGGCGGCCGCATCGTGC  3060

Qy  3061  GGGGGTCACACGTGGTGGGCCCGGTGGGCCCGGTGGTGATCGGCGGCGGGTGCCGTCGTCGCAACTCCA  3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  GGGGGTCACACGTGGTGGGCCCGGTGGGCCCGGTGGTGATCGGCGGCGGGTGCCGTCGTCGCAACTCCA  3120
```

Right column:

```
Qy  3121  GTGTCGGCCCGTACACCTTCCATCGGGGACTGCGGGACTGCGGGTCGCCGGTCGAGCT  3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121  GTGTCGGCCCGTACACCTTCCATCGGGGACTGCGGGACTGCGGGTCGCCGGTCGAGCT  3180

Qy  3181  ACTCCGTCCTGCGCGGCGCGCGCGTCGAGGGGCGTCCCCGCGCCAGGCTCACCGGCTGGTGA  3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  ACTCCGTCCTGCGCGGCGCGCGCGTCGAGGGGCGTCCCCTCA  3240

Qy  3241  TCGGGCCGGCGGTCGTCGGCGCCGTCTCCCGCAGGCTCACCGACTGGTGTA  3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241  TCGCCGGCGGCGGTCGTCGGCGCCGTCGTCCCGCAGGCTCACCGACTGGTGA  3300

Qy  3301  TCGGCGACCACAGCAAGGTGTATCTCACCACCACGACGACCATCCTGTCACCCGCGG  3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  TCGGCGACCACAGCAAGGTGTATCTCACCACCACGACGACCATCCTGTCACCCGCGG  3360

Qy  3361  AGCCGGCTTCATTCGCTCCGCCTACGTCCCGCGGCTCCTTCGCCCGGGCCCCCGGCGCG  3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361  AGCCGGCTTCATTCGCTCCGCCTACGTCCCGCGGCTCCTTCGCCCGGGCCCCCGGCGCG  3420

Qy  3421  CGTCGGCGGTGACCGTCCTCGACAAAACTCACCTACGCCCGCAGCCTCGGCACGC  3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421  CGTCGGCGGTGACCGTCCTCGACAAAACTCACCTACGCCCGCAGCCTCGGCACGC  3480

Qy  3481  GGTGCGTGACCATCCCGGCGGCCTCACCTTCGTCCCAGGGCGACGTGCTGCCACCGGCTCGT  3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  GGTGCGTGACCATCCCGGCGGCCTCACCTTCGTCCCAGGGCGACGTGCTGCCACCGGCTCGT  3540

Qy  3541  CGACACGCTGGCCGGCCGCTTCGGGTTACGCATCACCTCACTCGTGCGGCCGAGTCGCACGTCGA  3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  CGACACGCTGGCCGGCCGCTTCGGGTTACGCATCACCTCACTCGTGCGGCCGAGTCGCACGTCGA  3600

Qy  3601  CCGCTCATCACCGACACAGCGGTCCTTCCTGGGGCGTGCCTGGCACCGCGCCAGGTCCT  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  CCGCTCCATCACCGACACCAGCGGTCCTTCCTGGGGCGTGCCTGGCACCGCGCCAGGTCCT  3660

Qy  3661  GCTCGACGCGTGGCCACGGTGCCACTTCGTGCGACGTCTCCACCGACGAGT  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  GCTCGACGCGTGGCCACGGTGCCACTTCGTGCGACGTCTCCACCGACGAGT  3720

Qy  3721  GTACGGCTCCTCCCGCACGGGGCGCCGGAGAGCGACCCCCGCTTCGGACCTCGCC  3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  GTACGGCTCCTCCCGCACGGGGCGCCGGAGAGCGACCCCCGCTTCGGACCTCGCC  3780

Qy  3781  GTACGCGGCGCACGTCGAAGGGCCTCATGGCGTCGACGACGTGCCACCCACCG  3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781  GTACGCGGCGCACGTCGAAGGGCCTCATGGCGTCGACGACGTGCCACCCACCG  3840

Qy  3841  CCTTGGACGTCCGGGTGACCCGTCGTTGAACAACTTCGGCGGCGCCATCCCCGAGA  3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841  CCTTGGACGTCCGGGTGACCCGTCGTTGAACAACTTCGGCGGCGCCATCCCCGAGA  3900

Qy  3901  GCTCATACCGGCGCTTCCTGACCAGCCTCTGTCCCCTCTACGGCGA  3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901  GCTCATACCGGCGCTTCCTGACCAGCCTCTGTCCCCTCTACGGCGA  3960

Qy  3961  CGGGCGGCGCACGTGCGGGACCTGCGGACGACGTCAGGGGCCGTCGAACTCGT  4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961  CGGGCGGCGCACGTGCGGGACCTGCGGACGACGTCAGGGGCCGTCGAACTCGT  4020

Qy  4021  CCCGGTGTCGGGGGTCACACGGTTGCTCGGCGGGAGATTACAACATCGGGGCGCGGCCCAA  4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021  CCCGGTGTCGGGGGTCACACGGTTGCTCGGCGGGAGATTACAACATCGGGGCGCGGCCCAA  4080

Qy  4081  CCTTGGAGGTCACGGCAGCACCGTTGCTCGCACTGGCGGCGGCCCGAGGGGCGCATCGTGC  4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081  CCTTGGAGGTCACGGCAGCACCGTTGCTCGCACTGGCGGCGGCCCGAGGGGCGCATCGTGC  4140

Qy  4141  CGTCGAGAACCGGCGTCGTACGGGGCACCAGCAGCAAGATCACCGC  4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4141  CGTCGAGAACCGGCGTCGTACGGGGCACCAGCAGCAAGATCACCGC  4200
```

Qy  4201  GGAACTCGGTTACCGGCCGCGCACCGACTTCGCGACCGCTGGCCGCGAAGTG  4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4201  GGAACTCGGTTACCGGCCGCGCACCGACTTCGCGACCGCTGGCCGCGAAGTG  4260

Qy  4261  GTACGAGGGCACGAGGACTCGGCGACCGTGCGTCCCCTGCTCCCGCGACATGACGTCGGCCCG  4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4261  GTACGAGGGCACGAGGACTCGGCGACCGTGCGTCCCCTGCTCCCGCGACATGACGTCGGCCCG  4320

Qy  4321  ACCGCAACCACCGGCCGCCCGGCCACCGCCCCCGGCCGTCCGGTCGGCCGGTCAG  4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4321  ACCGCAACCACCGGCCGCCCGGCCACCGCCCCCGGCCGTCCGGTCGGCCGGTCAG  4380

Qy  4381  CGTCCGTGAGCGCGGCCGGCCGGCGCCGCGCCGGGGGCGGCGTGGACCACC  4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4381  CGTCCGTGAGCGCGGCCGGCCGGCGCCGCGCCGGGGGCGGCGTGGACCACC  4440

Qy  4441  AGTTCCGGCATGAAGAAGACGAATTCGGTGCGTGCCGTTCCGCTCATCCTCCTCCAGC  4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441  AGTTCCGGCATGAAGAAGACGAATTCGGTGCGTGCCGTTCCGCTCATCCTCCTCCAGC  4500

Qy  4501  AGTGCGTCCACGGGCGACCTGCCCCATCGCTTGACGGCCTGTCTGATGGTGGTCAGGGGA  4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4501  AGTGCGTCCACGGGCGACCTGCCCCATCGCTTGACGGCCTGTCTGATGGTGGTCAGGGGA  4560

Qy  4561  GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCACCACCGAGATGTCACCGGGA  4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4561  GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCACCACCGAGATGTCACCGGGA  4620

Qy  4621  ACCGTGAGAACCCCGGCCGGCGCGGGGCCCCGCAGGGCCCATCATGTCCGCTGGCGG  4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4621  ACCGTGAGAACCCCGGCCGGCGCGGGGCCCCGCAGGGCCCATCATGTCCGCTGGCGG  4680

Qy  4681  CACATGACGGCGGTGCAGCCAGTCGATCAGCCGCGGCGGGCGCCTGGCCCCTCC  4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4681  CACATGACGGCGGTGCAGCCAGTCGATCAGCCGCGGCGGGCGCCTGGCCCCTCC  4740

Qy  4741  AGGGAGAACAGCCAGTCGTGCACGAGCTCCTGGACTCCGGCGCCGACACTCCCAGGTGC  4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4741  AGGGAGAACAGCCAGTCGTGCACGAGCTCCTGGACTCCGGCGCCGACACTCCCAGGTGC  4800

Qy  4801  TCCCGCACGCGGGAACGCCGGCCGCTCGATCTTCGCTCACCGCGAAGCGGGCGGGCGGC  4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4801  TCCCGCACGCGGGAACGCCGGCCGCTCGATCTTCGCTCACCGCGAAGCGGGCGGGCGGC  4860

Qy  4861  CCGAACGGGCGAGGCCGAGCGCGCTCGTCGTGCGCCCAGGTGCGCCACGGCCCAGGCGC  4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4861  CCGAACGGGCGAGGCCGAGCGCGCTCGTCGTGCGCCCAGGTGCGCCACGGCCCAGGCGC  4920

Qy  4921  ATCGGCGGTGTCCGGTCGTGTCCTCGGGGACGAAGGGTGCCTCGGGGCGAGAACCCGTTC  4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4921  ATCGGCGGTGTCCGGTCGTGTCCTCGGGGACGAAGGGTGCCTCGGGGCGAGAACCCGTTC  4980

Qy  4981  ACGAGGACGAAGGGCACCTGCGCGTGCGTGCAGCCGGCGCCGTACCGTCCGGTCTCGCGTG  5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4981  ACGAGGACGAAGGGCACCTGCGCGTGCGTGCAGCCGGCGCCGTACCGTCCGGTCTCGCGTG  5040

Qy  5041  GTGTCGACGCGTGAGCGGGTTGCTCCGGGGTCTTGCCTGCGGGGGAAGAAGGGGTTGTCC  5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5041  CCCTGACGCGTGAGCGGGTTGCTCCGGGGTCTTGCCTGCGGGGGAAGAAGGGGTTGTCC  5100

Qy  5101  TCCGTGAGTTCGTCGGGGGTGACCAGTCGCGGGTCGGGCTGCGTGCGGGTCGGGCGGGTGTAG  5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5101  TCCTGAGTTCGTCGGGGGTGACCAGTCGCGGGTCGGGCTGCGTGCGGGTCGGGCGGGTGTAG  5160

Qy  5161  CCCTGACGCGTGAGCGGGTGACCAGTCGGGGGAAGAAGGGGTTGTCC  5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5161  CCCTGAGCGCGTGAGCGGGTGACCAGTCGGGGGAAGAAGGGGTTGTCC  5220

Qy  5221  AGTTCGGGGGTGACCAGTCCGACCGTCGGCCGCGGGCCGTGTCGTGCGGCGGTGTAG  5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5221  AGTTCGGGGGTGACCAGTCGGACCGTCGGCCGCGGGCCGTGTCGTGCGGCGGTGTAG  5280

Qy  5281  CCCAGCGCGTCCAGTGCGGGGTCAGCACCGGTGGGGTGCGGGGTGCCCACACCGCGCGCCA  5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  5281  CCCAGCGCGTCCAGTGCGGGGTCAGCACCGGTGGGGTGCGGGGTGCCCACACCGCGCGCCA  5340

Qy  5341  CCGTTCAGCACCGGCTGACCGTGGCCTTGCTGCAGCCGAGTGCGGGTGCCGATGCGGCCG  5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5341  CCGTTCAGCACCGGCTGACCGTGGCCTTGCTGCAGCCGAGTGCGGGTGCCGATGCGGCCG  5400

Qy  5401  AGCCGCATGTCATGGCAACGCACTCTACCTGTCGGGGCGTGCCACCCGCGC  5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5401  AGCCGCATGTCATGGCAACGCACTCTACCTGTCGGGGCGTGCCACCCGCGC  5460

Qy  5461  GCGGAACCGGCGGACTGCGGGGGGCACGGCCGCCCGCGACCACGGCGCCCGAAA  5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5461  GCGGAACCGGCGGACTGCGGGGGGCACGGCCGCCCGCGACCACGGCGCCCGAAA  5520

Qy  5521  CGATGGCTGAAAATGCTTGCAGCAAATTGCCGCCAACGTCTTTCGCGGCCGGCTTTTCGATCCT  5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5521  CGATGGCTGAAAATGCTTGCAGCAAATTGCCGCCAACGTCTTTCGCGGCCGGCTTTTCGATCCT  5580

Qy  5581  GTTACGTTCCTGGCAACCCCGGCCGCGCAGAAGCGGTTGGCGTGAGGCGTCCAGACCT  5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5581  GTTACGTTCCTGGCAACCCCGGCCGCGCAGAAGCGGTTGGCGTGAGGCGTCCAGACCT  5640

Qy  5641  CCGCCCGATTCCGGGATCATCAGGGGAGTTCACAATGCCGGCGTGGCCACCG  5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5641  CCGCCCGATTCCGGGATCATCAGGGGAGTTCACAATGCCGGCGTGGCCACCG  5700

Qy  5701  CGCTGTTCGCGGCGTGTGGCCATGACGGCATCGCGTGTGGCGGGCGGCGACAACGGCGGAA  5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5701  CGCTGTTCGCGGCGTGTGGCCATGACGGCATCGCGTGTGGCGGGCGGCGACAACGGCGGAA  5760

Qy  5761  GCGGTACCGACGCGGCGGCCAGTGTCGTCGGGGACCGTCACCTTCTGGGACACGTCCA  5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5761  GCGGTACCGACGCGGCGGCCAGTGTCGTCGGGGACCGTCACCTTCTGGGACACGTCCA  5820

Qy  5821  ACGAAGCCGAGAAGGCGACGTACCAGGCCCTCGCCGAGAGGCGTTCGAGGAAGGAGCACCCGA  5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5821  ACGAAGCCGAGAAGGCGACGTACCAGGCCCTCGCCGAGAGGCGTTCGAGGAAGGAGCACCCGA  5880

Qy  5881  AGGTCGACGTCAAGTACGTCAACGTCCCCGTTCGGCGGAGGCGAACGCCAAGTTCAAGAACG  5940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5881  AGGTCGACGTCAAGTACGTCAACGTCCCCGTTCGGCGGAGGCGAACGCCAAGTTCAAGAACG  5940

Qy  5941  CCGCGGGCGGCAACTCCGGTGCGCCCCAGTGTGATGCGTGATCGGACGGCTCGCCTGCGGCGG  6000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5941  CCGCGGGCGGCAACTCCGGTGCGCCCCAGTGTGATGCGTGATCGGACGGCTCGCCTGCGGCGG  6000

Qy  6001  ACTTCCCAGCATCGGCTACTCCGGCCCGCTCGACGGCACGCCCCTCGACGGCGTTC  6060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6001  ACTTCCCAGCATCGGCTACTCCGGCCCGCTCGACGGCACGCCCCTCGACGGCGTTC  6060

Qy  6061  CGGACCACCTTCCCCCAGGGCGGCAGCACCAGGTGCGGGAAGAGACCTACGCGGTTCCCGC  6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6061  CGGACCACCTTCCCCCAGGGCGGCAGCACCAGGTGCGGGAAGAGACCTACGCGGTTCCCGC  6120

Qy  6121  AGGTGATCGACACCCTGCGGCCTCTTCTACAACAAGGAGAACTGCTACGACGAGTTCAAGAACG  6180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6121  AGGTGATCGACACCCTGCGGCCTCTTCTACAACAAGGAGAACTGCTACGACGAGTTCAAGAACG  6180

Qy  6181  AGGTGCCGGGCTCCCTGCGGTCGGGGTCTTGCGTGAGACGAGAGACCG  6240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6181  AGGTGCCGGGCTCCCTGCGGTCGGGGTCTTGCGTGAGACGAGAGACCG  6240

Qy  6241  GCGCGAGCGGCGGACCGTCTACTGCGGCAGTGCCGCGTACTTGGTTCTCGACGAGAGACCG  6300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6241  GCGCGAGCGGCGGACCGTCTACTGCGGCAGTGCCGCGTACTTGGTTCTCGACGAGAGACCG  6300

Qy  6301  GGGAGGGCGGCGGACCTGGTCGTGACGCTCGACGAGAAGAACAAGACCGTCACGGTCGACGACGAAGCCG  6360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6301  GGGAGGGCGGCGGACCTGGTCGTGACGCTCGACGAGAAGAACAAGACCGTCACGGTCGACGACGAAGCCG  6360

Qy  6361  GTGTCGCGCGCCTCACCGGTGGTGCACCGAGTCGCGGGTGCGCAGGCGCAAGGCGGCCATCACCGACG  6420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   6361  GTGTGCGCGCCTACCCGGTCATCATCAAGGACCTTCTGTGGACAGCAGCAAGGCGGCCATCACCGACG  6420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6421  CGTCCGACGGCTGGAACAACATGCAGAACGCCTTCAAGTCGGGCAAGTCGGCCATGATGG          6480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421  CGTCCGACGGCTGGAACAACATGCAGAACGCCTTCAAGTCGGGCAAGTCGGCCATGATGG          6480

Qy   6481  TCAACGGCCCCTGGGCCATCGAGGACGTCAAGGCGGGAGCCCGCTTCAAGGACGACCGGCA          6540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481  TCAACGGCCCCTGGGCCATCGAGGACGTCAAGGCGGGAGCCCGCTTCAAGGACGACCGCGA          6540

Qy   6541  ACCTGGGGTGCCCCGGCCGCGCGGCAGAGGCTCTCCCCAGGGCGGGT                      6600
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   6541  ACCTGGGGTGCCCCGGCCGCGCGGCAGAGGCTCTCCCCAGGGCGGGT                      6600

Qy   6601  GGAACCTTCGGTGTACGCGGGCTCGAAGAACCTCGACGCCTTCTGTGAAGT                  6660
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601  GGAACCTTCGGTGTACGCGGGCTCGAAGAACCTCGACGCCTTCTGTGAAGT                  6660

Qy   6661  ACATGAGCTCCGCCAAGGTGCAGCAGCAGACCACCGAGAAGCTGAGCCTGCTGCCCACCC        6720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661  ACATGAGCTCCGCCAAGGTGCAGCAGCAGACCACCGAGAAGCTGAGCCTGCTGCCCACCC        6720

Qy   6721  GCACGTCCGTCTACGAGGTCCGTCGCGGACAACAGAGATGGTGAAGTTCTTCAAGC            6780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721  GCACGTCCGTCTACGAGGTCCGTCGCGGACAACAGAGATGGTGAAGTTCTTCAAGC            6780

Qy   6781  CGGCCGTGCGACAAGGCCGTCGAACGCCGTGAACGGCGCCGTGAATGCCGAGGGCAATGCCCTTCGCAGC   6840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781  CGGCCGTGCGACAAGGCCGTCGAACGCCGTGAACGGCGCCGTGAATGCCGAGGGCAATGCCCTTCGCAGC   6840

Qy   6841  CGATCCGGCTGCAG  6854
           ||||||||||||||
Db   6841  CGATCCGGCTGCAG  6854
```

```
RESULT 2
AR174057
LOCUS        AR174057              6854 bp    DNA        linear    PAT 17-DEC-2001
DEFINITION   Sequence 7 from patent US 6306627.
ACCESSION    AR174057
VERSION      AR174057.1   GI:17914377
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 6854)
  AUTHORS    Decker,H.
  TITLE      Isolation of the biosynthesis genes for pseudo-oligosaccharides
             from streptomyces glaucescens GLA.O, and their use
  JOURNAL    Patent: US 6306627-A 7 23-OCT-2001;
FEATURES             Location/Qualifiers
     source          1..6854
                     /organism="unknown"
BASE COUNT     1018 a   2445 c   2445 g    946 t
ORIGIN

Query Match       100.0%;   Score 6854;   DB 6;   Length 6854;
Best Local Similarity  100.0%;   Pred. No. 0;
Matches 6854;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy      1  CTGCAGGGTTCCTGTGCACGACCCCGCCCCGTCCTGGTCGACACCGGCGCTGTCGCAGAT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CTGCAGGGTTCCTGTGCACGACCCCGCCCCGTCCTGGTCGACACCGGCGCTGTCGCAGAT  60

Qy     61  CGGCGGCGATGTCGGCGATGTCGTGGCTGTGGAGCACCGGCACCACCGGTGCCCTTGGACAGC  120
           ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     61  GCTTCCGGCGCTCTCGGCGATGTCGTGGCTGTGGAGCACCACCGGCACCACCGGTGCCCTTTGACAGC  120

Qy    121  ATCGCCAGTCCGGCGTTGACCAGCCGGCGCACCGGCTCCTTCAGCACCATGTCGAGGCCCGATCGTGG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  ATCGCCAGTCCGGCGTTGACCAGCCGGCGCACCGGCTCCTTCAGCACCATGTCGAGGCCCGATCGTGGG  180

Qy    181  CTCGTCCCAGAACAGCAGCACGACGGCCGGGTCGTGCAGCAGCTCGCCGGCGATCTCGGCGCGCAT  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    181  CTCGTCCCAGAACAGCAGCACGACGGCCGGGTCGTGCAGCAGCTCGCCGGCGATCTCGGCGCGCAT  240

Qy    241  GCGCTGTCCGAGGCTGAGCTGAGCCTGTGCCACCGCGGGGTGACCGCGCGTCGATGTCGAGGAG  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GCGCTGTCCGAGGCTGAGCTGAGCCTGTGCCACCGGGGGTGACCGCGCGTCGATGTCGAGGAG  300

Qy    301  GTCCCGGAACAGGGCGAGGTTGCCGCCGGTAGACCGGTCCGGGGATGCGTCGTAGATGCGGGCG  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GTCCCGGAACAGGGCGAGGTTGCCGCCGGTAGACCGGTCCGGGGATGCGTCGTAGATGCGGGCG  360

Qy    361  CAGGATGCGGAAGGAGTCGGGTACCAGAGTCCCACCAGAGTCGGCTGCCGCTGGCTGCTGGCCGAA  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  CAGGATGCGGAAGGAGTCGGGTACCAGAGTCCCACCAGAGTCGGCTGCCGCTGGCTGCTGGCCGAA  420

Qy    421  GACGACGCCGATCGTGCGGCGTTGCGCTGCCGGTGCCGGCGTAGGGCTCCAGCCCGGCCGAC  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  GACGACGCCGATCGTGCGGCGTTGCGCTGCCGGTGCCGGCGTAGGGCTCCAGCCCGGCCGAC  480

Qy    481  CGTGCAGCCGCCGGAGGTGGGGTCATGATGCCGTCGTAGCCGTCAGCATCTTGATGATCGTGGTCGACTT  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CGTGCAGCCGCCGGAGGTGGGGTCATGATGCCGTCGTAGCCGTCAGCATCTTGATGATCGTGGTCGACTT  540

Qy    541  GCCGGCTCCGTTGGCGCCGATGTAGGCGGTCTTCGTCGCCGTGTATCTCGAAGGAGAC  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GCCGGCTCCGTTGGCGCCGATGTAGGCGGTCTTCGTCGCCGTGTATCTCGAAGGAGAC  600

Qy    601  GTCGTCGACGGCGCGCACGACGCGCGGTACCGGCGGGTCAGGAGGTGGAGAGGCTGCCGAG  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GTCGTCGACGGCGCGCACGACGCGCGGTACCGGCGGGTCAGGAGGTGGAGAGGCTGCCGAG  660

Qy    661  CAGGCCGGGCTCGGTTCGGCGTTCGGCGAACTCCTTGACGAGGTGTTCGCCGACGATCAC  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CAGGCCGGGCTCGGTTCGGCGTTCGGCGAACTCCTTGACGAGGTGTTCGCCGACGATCAC  720

Qy    721  GCGATCACCCGCTCGACGGCCGTCTCGACGAGGCCGCAGGCGCCGTCTCCGACGGCCGCCTCG  780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  GCGATCACCCGCTCGACGGCCGTCTCGACGAGGCCGCAGGCGCCGTCTCCGACGGCCGCCTCG  780

Qy    781  TCGAGGGTGAACGGCGGTGCCAGCCGCAGGATGTGGCCCCCAGGGAGTGGCGCAGCCCC  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  TCGAGGGTGAACGGCGGTGCCAGCCGCAGGATGTGGCCCCCAGGGAGTGGCGCAGCCCC  840

Qy    841  AGGTCGAGGGCGGTGGTGTAGACGGCGGTCTCGGGGGCGTGCCCGGCCGACG  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  AGGTCGAGGGCGGTGGTGTAGACGGCGGTCTCGGGGGCGTGCCCGGCCGACG  900

Qy    901  GCGTCGGTGACGAACTCCAGGCCCACAGCAGTCCGAGGCGCCGCTACCTGGCCGAGCTGG  960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GCGTCGGTGACGAACTCCAGGCCCACAGCAGTCCGAGGCGCCGCTACCTGGCCGAGCTGG  960

Qy    961  GGGAAGCGGGACTCCAGGCGCGCAGCCGCTCCTGGATGAGCTCGCGGGCCGACACG  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  GGGAAGCGGGACTCCAGGCGCGCAGCCGCTCCTGGATGAGCTCGCGGGCCGACACG  1020

Qy   1021  CGGTCGATCAGCCGGTCGCCGACGCCTTCCAGCGTGGCGGGCGGGGATCCCC  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  CGGTCGATCAGCCGGTCGCCGACGCCTTCCAGCGTGGCGGGCGGGGATCCCC  1080

Qy   1081  AGTGGGTTGCTCGGTACGTCGAGGGCGTACGCCCCTCGGCCTGCA  1140
           |||||||||||||||||||||||||||||||||||||||||||||
Db   1081  AGTGGGTTGCTCGGTACGTCGAGGTCGGGTACGCCCCTCGGCCTGCGCA  1140

Qy   1141  GCTTCCGGCGCTCGGCGATCCGGCGCAGCCCAGGGAATCCGCTCGCGGTGCCCTTGGACAGC  1200
           ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1141  GCTTCCGGCGCTCCGGCGATCCGCGCCAGCCAGGGGCGGCGGAATCCGCTCGCGGTGCCCTTTGACAGC  1200

Qy   1201  ATCGCCAGTCCGGCGTTGACCAGCCGGCGCACCGGCTCCTTCAGCACCATGTCGAGGCCCGATCGTGGG  1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  ATCGCCAGTCCGGCGTTGACCAGCCGGCGCACCGGCTCCTTCAGCACCATGTCGAGGCCCGATCGTGGG  1260

Qy   1261  CCGGCGTGAGGACCTCGTCGCGGACGACAGCAGCACGACGGCCGGGTCGTGCAGCAGCACGCCGGCCGGCAT  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1261 CCGCCGGTGAGGACCTCGTCGCGACGAGCAGCACGCCGCCGCCGTCCCGGCAGGCGCCGGCG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 ATCCGCTCCCAGTAGCCGGGGGGGCCGGCCACGATGACGCGGGGGGCCGCCGGAGGACGGGTTCG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ATCCGCTCCCAGTAGCCGGGGGGGCCGGCCACGATGACGCGCTGCCGGCCGGAGGACGGGTTCG 1380

Qy   1381 AAGACCAGGGCCGAGACGTTGGGCTTCTCCGCGATGTGCCGGCGCACGAGGGTCGCGCAC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAGACCAGGGCCGAGACGTTGGGCTTCTCCGCGATGTGCCGGCGCACGAGGGTCGCGCAC 1440

Qy   1441 CGCACGTCGCCACGAGGGGTACTCCAGGCCCAGGGGACAGCAGCGGTAGCCAGTAGGGGCTGTA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CGCACGTCGCCACGAGGGGTACTCCAGGCCCAGGGGACAGCAGCGGTAGCCAGTAGGGGCTGTA 1500

Qy   1501 GCCAGCACGCTGTTGCGCGTGAAGGCCTGGGCCGAGATGTCCCAGTGGACCAGCATCCGG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCAGCACGCTGTTGCGCGTGAAGGCCTGGGCCGAGATCCCCAGTGGACCAGCATCCGG 1560

Qy   1561 GCGCCCATGGTGCTGCCGTGGACGACCCCGAGGGGGCCGCAGGGGCGCAGATCCGGCGCGCCGGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GCGCCCATGGTGCTGCCGTGGACGACCCCGAGGGGGCCGCAGGGGCGCAGATCCGGCGCGCCGGC 1620

Qy   1621 GCGGGGGTGCCCTGGACGACCCGCTCGACCACCTCGACCACCTCGCGCCGGTTGCGCCGCCGGC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCGGGGGTGCCCTGGACGACCCGCTCGACCACCTCGACCACCTCGCGCCGGTTGCGCCGCCGGC 1680

Qy   1681 AAGGCTAGGTGTGTCGGAGCTGTTCGGGCAGCAGCCTGGCGAGCAGTTCCAGGCCGGCGGCG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AAGGCTAGGTGTGTCGGAGCTGTTCGGGCAGCAGCCTGGCGAGCAGTTCCAGGCCGGCGGCG 1740

Qy   1741 CGGTTCCGGCGTGTCGTGGCGCTGTGCCGGTGCCCTGGACGTTCCACAGGCGGCGGTGAGTGCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CGGTTCCGGCGTGTCGTGGCGCTGTGCCGGTGCCCTGGACGTTCCACAGGCGGCGGTGAGTGCC 1800

Qy   1801 TCGACGACCTCCGGGTGCCCGGTGCCCGTGGGTGGTGACTGGGTGAGGGTTCCCGGCGAAGTCG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TCGACGACCTCCGGGTGCCCGGTGCCCGTGGGTGGTGACTGGGTGAGGGTTCCCGGCGAAGTCG 1860

Qy   1861 AGGTACTGGTTGCCGTCAGGTCGTTGCGGCGTGTCAGAACGGGACCCGCGTCTCCGGCGAAGACCCCGG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGGTACTGGTTGCCGTCAGGTCGTTGCGGCGTGTCAGAACGGGACCCGCGTCTCCGGCGAAGACCCCGG 1920

Qy   1921 CGTCCGTGGACGGCGTTCCTCGGAGGGCGCCGGCCCCGCCAGTTGGCGGGGTTGGCGGGGGCCTTCCCGTGCCAGG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CGTCCGTGGACGGCGTTCCTCGGAGGGCGCCGGCCCCGCCAGTTGGCGGGGTTGGCGGGGGCCTTCCCGTGCCAGG 1980

Qy   1981 TGCTGTGTCTGGTTGACGTCTTAAGGCTGTCATCGCGTCGTCGGCGTCGGCCTCGGCGTGACGCGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TGCTGTGTCTGGTTGACGTCTTAAGGCTGTCATCGCGTCGTCGGCGTCGGCCTCGGCGTGACGCGAT 2040

Qy   2041 CGGCCGGAAACTGCGTTGGCGCACCGGTTGGGCGGTCGGCGCTGCGCGTCGAGTCAAACAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CGGCCGGAAACTGCGTTGGCGCACCGGTTGGGCGGTCGGCGCTGCGCGTCGAGTCAAACAC 2100

Qy   2101 TTGAACACACACCCTGCAAGAGTTGTTTGCGGCGCCACACCGGCCGTTCAGGAAGTTGCGAGCGGCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TTGAACACACACCCTGCAAGAGTTGTTTGCGGCGCCACACCGGCCGTTCAGGAAGTTGCGAGCGGCCC 2160

Qy   2161 CGGCCACTCTGGTTGAGTCGACGTGCTACGGCGCCACCACGCCTTCACGTTGGCGCCACGTCGAGGAGGGA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CGGCCACTCTGGTTGAGTCGACGTGCTTACGGCGCCACCACGCCTTCACGTTCGAGGAGGGA 2220

Qy   2221 CCTGTGAGAACAAGCCCAGCAGCAAGGTGATCTCGCCGAGGGCCGAGTGAAGGCCCTGG 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CCTGTGAGAACAAGCCCAGCAGCAAGGTGATCTCCACCCCGAGGGCGAGTGAAGGCCCTG 2280

Qy   2281 TCCTGGCAGGTGAACCGTGGAACCGGCACGAGAACTGAGGCCGGCCACGCCCAAGCAGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCCTGGCAGGTGAACCGTGGAACCGGCACGAGAACTGAGGCCGGCCACGCCCAAGCAGC 2340

Qy   2341 TGCTCCCCATCGCCAAACAGCAGCCCGTGCCTGCTCTTCTTACGCGCTGGAGTCCCCTGCGCGCG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TGCTCCGGTGTGACCGTGCTCCTGCAACAAGTCACCTCACCTGGAGTCCCTGCGCGCG 2400
```

```
Qy   2401 GTGTCCGGGAGGCCGGCGGCGTCGTCGTGGGCGGTACGCGGGGAGATCCGGGAACTCACCG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GTGTCCGGGAGGCCGGCGGCGTCGTCGTGGGCGGTACGCGGGGAGATCCGGGAACTCACCG 2460

Qy   2461 GCGACGGCACCGCGTTCGGGGTTACGCATCACCTTCCACCAGCCCCGCCGGCTCCGGTC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GCGACGGCACCGCGTTCGGGTTACGCATCACCTTCCACCAGCCCCGCCGGCTCGGTC 2520

Qy   2521 TCGGCCGCACGGCGTGCCATCGCCCGGGCTTCCTGGGCGACGACGACTTCCTGCTGTACC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TCGGCCGCACGGCGTGCCATCGCCCGGGCTTCCTGGGCGACGACGACTTCCTGCTGTACC 2580

Qy   2581 TGGGGACAACTACCTGCCCCGCGGCGTCACCCGGCGCTTCGCCCGCACTTCGCCCGCCGGATC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TGGGGACAACTACCTGCCCCGCGGCGTCACCCGGCGCTTCGCCCGCACTTCGCCCGCCGGATC 2640

Qy   2641 CCGCGGCGCCCGGCTGCTCGCTGCCGCTGCGCCGAACCCGTGCGCCACCCGTGCGCCTCCGGCCAGCT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CCGCGGCGCCCGGCTGCTCGCTGCCGCTGCGCCGAACCCGTGCGCCACCCGTGCGCCTCCGGCCAGCT 2700

Qy   2701 AGGTCCACGCGGACGGGAACGTGCTCGGCGTGTACGCCGTGTACGCCGTCCAGCCCGGGCCA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 AGGTCGACGCGGACGGGAACGTGCTCGGCGTGTACGCCGTGTACGCCGTCCAGCCCGGGCCA 2760

Qy   2761 CGCTCGCGCTGCTCATCGCCCGCGGCGAGCTGGAGATCACCCGCCGTGCAGTCGGATGATCGACC 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 CGCTCGCGCTGCTCATCGCCCGCGGCGAGCTGGAGATCACCCGCCGTGCAGTCGGATGATCGACC 2820

Qy   2821 TCACCCCCTCCGCCCGCGGCGAGCTGGAGATCAACCACCACCGGCCCTGGCGCGACCAGCGCGCGG 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 TCACCCCCTCCGCCCGCGGCGAGCTGGAGATCAACCACCACCGGCCCTGGCGCGACCAGCGCGCGG 2880

Qy   2881 GGGGCCTGCGCGTACGGCGGCCTGGCGTCAACCGTCAACGTCTCTGACGCGTCGGCCGATGGGGA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GGGGCCTGCGCGTACGGCGGCCTGGCGTCAACCGTCAACGTCTCTGACGTCCGGCCGATGGGGA 2940

Qy   2941 AGGACATGCTGAGGTCAACCGTGGTCGCGCTCACGGTTCGGATGGCCGATCGAGGGGA 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 AGGACATGCTGAGGTCAACCGTGGTCGCGCTCACGGTTCGGATGGCCGATCGAGGGGA 3000

Qy   3001 AGGTCGACGCGCACAGCAGCGTCGGTGGTCGCGCCCAGTGGCCGATCGTGCCGAGGGCCGCATCGTGC 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AGGTCGACGCGCACAGCAGCGTCGGTGGTCGCGCCCAGTGGCCGATCGTGCCGAGGGCCGCATCGTGC 3060

Qy   3061 GGGGGTCACACGTGGTTGGGCCCGGTGATCGGGCGGTGTCGCGGTGCCGTGTCGCGTCAGCAACTCCA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GGGGGTCACACGTGGTGGGCCCGGTGTGTGATCGGGCGGTGCCGTCGCGTCAGCAACTCCA 3120

Qy   3121 GTGTCGGCCGCGTACACCTTCCATCGGGCGTACACCGTGTATCTCACCCGGCGCCGCCGTCACCGGCGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GTGTCGGCCGACCACGTTCCATCGGGCGTACACCGTGTATCTCACCCGGCGCCGCCGTCACCGGCGG 3180

Qy   3181 ACTCCGTCGTCCTGCGCGCGGGCCCCCCCAGTGCGCCATCGAGGGGTGCGTGC 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 ACTCCGTCGTCCTGCGCGCGGGCCCCCCCAGTGCGCCATCGAGGGGTGCGTGC 3240

Qy   3241 TCGGGCCGGCGGCCCGTCGTCCCGGCCCCCCGTCTCCCGGCGGCTCCCCAGGCTCACCGACTGTGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TCGGGCCGGCGGCCCGTCGTCCCGGCCCCCCGTCTCCCGGCGGCTCCCCAGGCTCACCGACTGTGA 3300

Qy   3301 TCGGGCGACCCAGCAAGGTGTATCTCGTCGTTCCGTCACCGGCGG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 TCGGGCGACCCAGCAAGGTGTATCTCCATCTCGTCACCGGCGG 3360

Qy   3361 AGCGGGCTTCATTCGGCTCCGCGCCCCTCCGCGCGCCCCCCGGCGCG 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 AGCGGGCTTCATTCGGCTCCGCGCCCCTCCGTGCGCCCCCGGCGCG 3420

Qy   3421 CGTCGCGGTGACCGTCCTCGACAAACTCACCTCACGCCGGCCAGCCTGCACGC 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CGTCGCGGTGACCGTCCTCGACAAACTCACCTCACGCCGGCCAGCCTGCACGC 3480
```

```
Qy  3481  GGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTCGT  3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  GGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTCGT  3540

Qy  3541  CGACACGCTGGCCGCGCGGCACGACGACATCGTCGGCCGCACTTCGCGGCCGAGTCGCACGTCGA  3600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  CGACACGCTGGCCGCGCGGCACGACGACATCGTCGGCCGCACTTCGCGGCCGAGTCGCACGTCGA  3600

Qy  3601  CCGCTCCATCACGGACAGCCGTGCCTTCACCCGCACCAACGTGCTGGCCACCCAGTTCCT  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  CCGCTCCATCACGGACAGCCGTGCCTTCACCCGCACCAACGTGCTGGCCACCCAGTTCCT  3660

Qy  3661  GCTCGACGCCGCGCTCCGCCACGTGTGCGCCACGTCTCGCACGTCTCCACCGACGAGGT  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  GCTCCGACGCCGCGCTCCGCCACGTGTGCGCCACGTCTCGCACGTCTCCACCGACGAGGT  3720

Qy  3721  GTACGGCTGCCTCCCCGCCAGGGCGCTCCTCCCCGCGAGAGCGACCCCCGTCTTCCGACCTCGCC  3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  GTACGGCTGCCTCCCCGCCAGGGCGCTCCTCCCCGCGAGAGCGACCCCCGTCTTCCGACCTCGCC  3780

Qy  3781  GTACCCGGCGCTCGAAGGCGGCCTCGGACCTCATGGCGCTCGCCCACCACCCCACGG  3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781  GTACCCGGCGCTCGAAGGCGGCCTCGGACCTCATGGCGCTCGCCCACCACCCCACGG  3840

Qy  3841  CCTCGACGTCCGGGTGACCCGCTGTTCGAACAACTTCGGCCCGCCCCACCAGCATCCCGAGAA  3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841  CCTCGACGTCCGGGTGACCCGCTGTTCGAACAACTTCGGCCCGCCCCACCAGCATCCCGAGAA  3900

Qy  3901  GCTCATACCGCGCTTCCTGACCACGCCGTCGGCGGCGACTGGCTCGCACGTCGCGACGTCGAACTCGT  3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901  GCTCATACCGCGCTCTTCCTGACCACGCCGTCGGCGGCGACTGGCTCGCACGTCGCGACGTCGAACTCGT  3960

Qy  3961  CGGGCGGCACGTGCGGCGACTGGCTGCGCGGACCGGCACGACGGCTACAAACATCGGCGCGCCCGGTAG  4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961  CGGGCGGCACGTGCGGCGACTGGCTGCGCGGACCGGCACGACGGCTACAAACATCGGCGCGCCCGGTAG  4020

Qy  4021  CGGCGTGTCGGGCCGGCCGCGGGAGAGATCTACAAACATCGGCGCGCGCCGCCCAA  4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021  CGCGCGTGTCGGGCCGGCCGCGGGAGAGATCTACAAACATCGGCGCGCGCCGCCCAA  4080

Qy  4081  CCTGAGCTCACGCACCGGTTGCTCCGCACTGTGCGGGCGCCCGGAGCGCATGTCCA  4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081  CCTGAGCTCACGCACCGGTTGCTCCGCACTGTGCGGGCGCCCGGAGCGCATGTCCA  4140

Qy  4141  CGTCGAGAACCGGCAGGGCACGACGGGCTACGCGGTCGACCACAGATCCAAGATCACCGGC  4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4141  CGTCGAGAACCGGCAGGGCACGACGGGCTACGCGGTCGACCACAGATCCAAGATCACCGGC  4200

Qy  4201  GGAACTCGGTTACCGGCGGGCCGGCACCGCACCGGCGACCCGCTGGCCGACACCGCGGAAGTG  4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4201  GGAACTCGGTTACCGGCGGGCCGGCACCGCACCGGCGACCCGCTGGCCGACACCGCGGAAGTG  4260

Qy  4261  GTACGAGGCGGCACGAGGACTGGTCGGGCGCGCGCGCGACATGACGTCGGGCCGG  4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4261  GTACGAGGCGGCACGAGGACTGGTCGGGCGCGCGCGCGACATGACGTCGGGCCGG  4320

Qy  4321  ACCGCAACCACCGCCCGCGCCCCGGCACCGGCGCGCGCGGCCGGTCAG  4380
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  4321  ACCGCAACCACCGCCCGCGCCCCGGCACCGGCGCGCGCGGCCGGTCAG  4380

Qy  4381  CGTCCGTGAGCCGGCGACTTCGGACCCGGCTGGCCGCGACACCGGAGTGGTCGGGGGA  4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4381  CGTCCGTGAGCCGGCGACTTCGGACCCGGCTGGCCGCGACACCGGAGTGGTCGGGGGA  4440

Qy  4441  AGTTCCGGCATGAAGACGAATTCGGTGCGCGGCGGCGGCGCGGCGAATTGCGCGACTCCTCCAGC  4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441  AGTTCCGGCATGAAGACGAATTCGGTGCGCGGCGGCGGCGCGGCGAATTGCGCGACTCCTCCAGC  4500

Qy  4501  CGTCGGTCCACGGCGTCCAACCCGGCGCGCCCGCCGGGGGCGCGGCGGGTCAGGGGGA  4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4501  CGTCGGTCCACGGCGTCCAACCCGGCGCGCCCGCCGGGGGCGCGGCGGGTCAGGGGGA  4560

Qy  4561  GGGTCGGTGAAGGCCATGAGCGGCCAGTCGTCGAAGCCGAGATCGTCGACCACCCGGGGA  4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4561  GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCGAGATGTCACCGGGGA  4620

Qy  4621  ACCGTGAGACCCGCGCGGCGCGGCCCCGCACGGCGCGCGGGCCATCATGTCGCTGCGG  4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4621  ACCGTGAGACCCGCGCGGCGCGGCCCCGCACGGCGCGCGGGCCATCATGTCGCTGCGG  4680

Qy  4681  CACATGACGCGCTGGTGTCAGCGCCAGGTCGATCGATCAGGAGCTCCTGCGACACAGG  4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4681  CACATGACGCGCTGGTGTCAGCGCCAGGTCGATCGATCAGGAGCTCCTGCGACACAGG  4740

Qy  4741  AGGGAGAACAGCGAGTGCTGCACGAGTCCTGCGACGAGCTCCGGAGTCGGGCGAGAACCCGTC  4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4741  AGGGAGAACAGCGAGTGCTGCACGAGTCCTGCGACGAGCTCCGGAGTCGGGCGAGAACCCGTC  4800

Qy  4801  TCCCGCACGCGCCCGGAACCCCTCGATCTTCCGCTGCACCGCGCACGGGGCGCGGGC  4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4801  TCCCGCACGCGCCCGGAACCCCTCGATCTTCCGCTGCACCGCGCACGGGGCGCGGGC  4860

Qy  4861  CGACGGCGAGGCCGACGCGCTTCGTCGCGCTCGGGGGAGACGAAGATGATGCCGAGCATC  4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4861  CGACGGCGAGGCCGACGCGCTCGTCGCGCTCGGGGGAGACGAAGATGATGCCGAGCATC  4920

Qy  4921  ATCGCGGCCCGGTCGTCCTCGTGGGAGACGAAGGGTCCTCGGGGCTGCGCGGGCGAACCCGTC  4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4921  ATCGCGGCCCGGTCGTCCTCGTCGGGAGACGAAGGGTCCTCGGGGCTGCGCGGGCGAACCCGTC  4980

Qy  4981  ACGAGGACGAAGGCACCTGCCGCTCGGTGCAGCGCCGTCCGGCGTACCGTCTCCGGCGGTG  5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4981  ACGAGGACGAAGGCACCTGCCGCTCGGTGCAGCGCCGTCGTGCTACCGTCTCCGGCGGTG  5040

Qy  5041  GTGTCCCGCGTGCAGTCGCGAGTCCGGAGACGACGAAGAGCATCGGGCGTCTCGCGGTGGCGAGCATC  5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5041  GTGTCCCGCGTGCAGTCGCGAGTCCGGAGACGACGAAGAGCATCGGGCGTCTCGCGGTGGCGAGCATC  5100

Qy  5101  TCCGTGAGTTCGTCCTGCGTGCCTGCTTCACCGGGCGTCGCGGGGCTGGGGCTGCGCGTGTGTTCC  5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5101  TCCGTGAGTTCGTCCTGCGTGCCTGCTTCACCGGGCGTCGCGGGGCTGGGGCTGCGCGTGTGTTCC  5160

Qy  5161  CCTGACCGTGAGCGCTGCCCCATCACCTGGGCCCATCACCTCGGGCGCGGGCGCGTCCTCGTAG  5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5161  CCTGACCGTGAGCGCTGCCCCATCACCTGGGCCCATCACCTCGGGCGCGGGCGCGTCCTCGTAG  5220

Qy  5221  AGTTCGGGGGTGACCAGTCGGCTCAGCGCGGTCATGGCGCAACGCACTCTACCTGTGCGGCCCGTCGTAG  5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5221  AGTTCGGGGGTGACCAGTCGGCTCAGCGCGGTCATGGCGCAACGCACTCTACCTGTGCGGCCCGTCGTAG  5280

Qy  5281  CCAGCCGGCTCCAGTGCGGTCAGCAGTGCCGGGTCACCGGGTCGCCGCCCGCCACCCGCCGCA  5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5281  CCAGCCGGCTCCAGTGCGGTCAGCAGTGCCGGGTCACCGGGTCGCCGCCCGCCACCCGCCGCA  5340

Qy  5341  CCGTTCAGCACCGGGTACCGGCTGTGCCTTGCTGACCGGGGCTGCGATGTCGGGGG  5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5341  CCGTTCAGCACCGGGTACCGGCTGTGCCTTGCTGACCGGGGCTGCGATGTCGGGGG  5400

Qy  5401  AGCCGCATGGTCATGGCGCAACGCACTCGTGTCGGGGGCGTGCCACCGGGGCCCACCGGGC  5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5401  AGCCGCATGGTCATGGCGCAACGCACTCGTGTCGGGGGCGTGCCACCGGGGCCCACCGGGC  5460

Qy  5461  GCGGAACCCGGCGACTGCGGGGACGGTTGCGCGGGCAGAGCCGGTTGGCCTGAGGCGTCCAGACCTT  5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5461  GCGGAACCCGGCGACTGCGGGGACGGTTGCGCGGGCAGAGCCGGTTGGCCTGAGGCGTCCAGACCTT  5520

Qy  5521  CGATGGCTGAAAATGCTTGCAGCAAATTCCGCAACCCGCCCGAAATTGCCGCAAATTGCCGCGTCTTTCGCGGTTTGATCCT  5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5521  CGATGGCTGAAAATGCTTGCAGCAAATTCCGCAACCCGCCCGAAATTGCCGCAAATTGCCGCGTCTTTCGCGGTTTGATCCT  5580

Qy  5581  GTTACGTTCCTGCGCAACCCGCCCGAGAGCCGGTTGGCGTGAGGCGTCCAGACCTT  5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5581  GTTACGTTCCTGCGCAACCCGCCCGAGAGCCGGTTGGCGTGAGGCGTCCAGACCTT  5640

Qy  5641  CCCCGGATTCCGGGGATCACTCAGGGAGTTCACAATGCGGCGGTTGCCGGCCACCG  5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   5641 CCGGCCGATTCCGGATCACTCAGGGGAGTTCACAATGCGGCGTGGCATTGCGGCCACCG 5700

Qy   5701 CGGTGTTCCGCGGCTGTGGCCATGACGGCATCGGCCGTGCGGGGGCGACAACGCGGAA 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 CGGTGTTCCGCGGCTGTGGCCATGACGGCATCGGCCGTGCGGGGGCGACAACGCGGAA 5760

Qy   5761 GCCGTACCGACGGGGCGGCCGGCACCGAGCTGTCGGGGACCGTCACCTTCTGGGACACGTCCA 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 GCCGTACCGACGGGGCGGCCGGCACCGAGCTGTCGGGGACCGTCACCTTCTGGGACACGTCCA 5820

Qy   5821 ACGAAGCCGAGAAGGCGACGACGTACCAGGCCGCTTCGAGAAGGAGGAGCACCCCGA 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 ACGAAGCCGAGAAGGCGACGACGTACCAGGCCGCTTCGAGAAGGAGGAGCACCCCGA 5880

Qy   5881 AGGTCGACGTCAAGTACGTCAACGTCCCCGTTCGGCGGAGCGGAACGCCAAGTTCAAGAACG 5940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5881 AGGTCGACGTCAAGTACGTCAACGTCCCCGTTCGGCGGAGCGGAACGCCAAGTTCAAGAACG 5940

Qy   5941 CCCGGGGCGGCCAACTCCCGGTGCCCCGACGTGATGCGTACGGAGGTCGCCTGGGTCGCGG 6000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 CCCGGGGCGGCCAACTCCCGGTGCCCCGACGTGATGCGTACGGAGGTCGCCTGGGTCGCGG 6000

Qy   6001 ACTTCGCCAGCAGCATCGGCTACCTCGCCCCGCTCCCCCCTCGACGACGGGT 6060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 ACTTCGCCAGCAGCATCGGCTACCTCGCCCCGCTCCCCCCTCGACGACGGGT 6060

Qy   6061 CGGACCACCTTCCCCAGGCGGCAGCACCAGGTACGAGGGGAAGACCTACGCCGGTCCCCC 6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 CGGACCACCTTCCCCAGGCGGCAGCACCAGGTACGAGGGGAAGACCTACGCCGGTCCCCC 6120

Qy   6121 AGGTGATCGACACACCTGCGCTCTTCTACAACAAGGAACTGCTGACGAAGGCCGGTGTCG 6180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121 AGGTGATCGACACACCTGCGCTCTTCTACAACAAGGAACTGCTGACGAAGGCCGGTGTCG 6180

Qy   6181 AGGTGCCGGGCTCCCTCCGCCGAGCTGAAGACGGCCGCCGCCGAGATCACCGAGAAGAGACCG 6240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181 AGGTGCCGGGCTCCCTCCGCCGAGCTGAAGACGGCCGCCGCCGAGATCACCGAGAAGAGACCG 6240

Qy   6241 GCGCCGAGCGGCTCTACTGTCGGGGCGACGACACCCGTACTGGTTCCTGCCCTACCTCTACG 6300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241 GCGCCGAGCGGCTCTACTGTCGGGGCGACGACACCCGTACTGGTTCCTGCCCTACCTCTACG 6300

Qy   6301 GGGAGGGCGGCGGACCTGGTCGACGAGAACAAGAACAGACCTCACGGTCGACGACGAAGCCG 6360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301 GGGAGGGCGGCGGACCTGGTCGACGAGAACAAGAACAGACCTCACGGTCGACGACGAAGCCG 6360

Qy   6361 GTGTGCCGGGCCTACCGGCCGTCATCAAGGACCTCGTGGACAGCAAGAGGCGGCCATCACCGACG 6420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361 GTGTGCCGGGCCTACCGGCCGTCATCAAGGACCTCGTGGACAGCAAGAGGCGGCCATCACCGACG 6420

Qy   6421 CGTCCGACGGCTGGAACAACATGCAGAACGCCTTCAAGTCGGCCAAGGTCGGCCATGATGG 6480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421 CGTCCGACGGCTGGAACAACATGCAGAACGCCTTCAAGTCGGCCAAGGTCGGCCATGATGG 6480

Qy   6481 TCAACGGCCCTGGCCATCCGGCAGACGTCAAGGCGGGAGCCCCGCTTCAAGGACGCGCCGGCA 6540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 TCAACGGCCCTGGCCATCCGGCAGACGTCAAGGCGGGAGCCCCGCTTCAAGGACGCGCCGGCA 6540

Qy   6541 ACCTGGGGGTCGCCCGCCCCCGTCCCCGGCAGTCGGCCGGACAGGGGCTCTCCCCCAGGGCGGGT 6600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541 ACCTGGGGGTCGCCCGCCCCCGTCCCCGGCAGTCGGCCGGACAGGGGCTCTCCCCCAGGGCGGGT 6600

Qy   6601 GGAACCTCTCGGTGTACGCGGGCTCGAAGAACCTCGACGCCTCGACGACGAACCTTCGTGAAGT 6660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601 GGAACCTCTCGGTGTACGCGGGCTCGAAGAACCTCGACGCCTCGACGACGAACCTTCGTGAAGT 6660

Qy   6661 ACATGAGCTCCGCCAAGGTGCAGCAGCAGAGACCACCGAGAAGCTGAGCCTGCTGCCCACCC 6720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 ACATGAGCTCCGCCAAGGTGCAGCAGCAGAGACCACCGAGAAGCTGAGCCTGCTGCCCACCC 6720

Qy   6721 GCACGTCCGTCTACGAGCTCTACGAGGTCCCCGTCCCGCGCCGACAACGAGAGTTCTTCAAGC 6780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 GCACGTCCGTCTACGAGCTCTACGAGGTCCCCGTCCCGCGCCGACAACGAGAGTTCTTCAAGC 6780

Qy   6781 CGGCCGTCGACAAGGCCGTCGAACGGCCGTGAACGCCGTGGATCGGCCGAGGGCAATGCCCTTCTTCGAGC 6840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 CGGCCGTCGACAAGGCCGTCGAACGGCCGTGAACGCCGTGGATCGGCCGAGGGCAATGCCCTTCTTCGAGC 6840

Qy   6841 CGATCCGGCTGCAG 6854
          ||||||||||||||
Db   6841 CGATCCGGCTGCAG 6854
```

```
RESULT 3
AF237894S1
LOCUS       AF237894S1              9523 bp    DNA     linear   BCT 26-MAR-2000
DEFINITION  Streptomyces antibioticus Tu99 oleandomycin biosynthetic gene
            cluster, partial sequence.
ACCESSION   AF237894
VERSION     AF237894.1  GI:7329189
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Streptomyces antibioticus
  ORGANISM  Streptomyces antibioticus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 9523)
  AUTHORS   Draeger,G., Park,S.-H.H. and Floss,H.G.
  TITLE     Mechanism of the 2-deoxygenation step in the biosynthesis of the
            deoxyhexose moieties of the antibiotics granaticin and oleandomycin
  JOURNAL   J. Am. Chem. Soc. 121, 2611-2612 (1999)
REFERENCE   2  (bases 1 to 9523)
  AUTHORS   Park,S.-H.H., Sohng,J.-K.K., August,P.R., Niggemann,J. and
            Floss,H.G.
  TITLE     A cluster of genes from Streptomyces antibioticus involved in the
            biosynthesis of the deoxysugar moieties of oleandomycin
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 9523)
  AUTHORS   Park,S.-H.H., Sohng,J.-K.K., August,P.R. and Floss,H.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2000) Chemistry, University of Washington, Box
            351700, Seattle, WA 98195-1700, USA
FEATURES             Location/Qualifiers
     source          1..9523
                     /organism="Streptomyces antibioticus"
                     /strain="Tu99"
                     /db_xref="taxon:1890"
     CDS             complement(167..1153)
                     /note="Orfil; involved in the 2-deoxygenation step in
                     dTDP-L-oleandrose biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-3,4-diketo-2,6-dideoxyglucose
                     3-ketoreductase"
                     /protein_id="AAF59931.1"
                     /db_xref="GI:7329191"
                     /translation="MPSPRLRFGVLGAADIALRRTVPALLAHPDVTVAVSSRDTARA
                     ARFAAAFGCEAVPGHQALLDRDDIDALYVPLPVMVHTPWVEAALLRGRHYLVERKPLTA
                     TRSGAEDLIALARSRGLVLMENFTSLHHAQHGTVTDLLRDGTIGELRSLSAAFTIPPK
                     PEGDIRYQPDVGGGALLDIGIYPLRAALHFLGPDLHAAGAVLRRERRNVVVSGHVLL
                     TTPHGVVAELAFGMEHAYRSEYTLFGTAGRLRLRDRAFTPPETHRPVEIHRQDALDIV
                     DLPPDAQFANLVRDEVLAVREGPGRLTQHHADAVRQADLVERVMAVARVRWC"
     CDS             complement(1173..2597)
                     /note="Orfl0; involved in the 2-deoxygenation step in
                     dTDP-L-oleandrose biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-4-keto-6-deoxyglucose 2,3-dehydratase"
                     /protein_id="AAF59932.1"
                     /db_xref="GI:7329192"
                     /translation="MIWGIPAMSEAMGSVPTAGSEVSSTCALLSWLDARRRANRLTVE
                     HVPFRELSGWQFDENTGNLRHTSGRFFSIEGLRVRTDHCWFGSWTQPIIVQPEIGILG
                     LLVKRFDGILHVLVQAKNEPGNIGGLQLSPTVQATRSNYTRVHRGGGVRYLEYFASPR
                     GRGRVLADVLQSEQGSWFLHKRNRNMEVEALDVPLDDFHWISLGGLRKLLLRPHLV
                     NMDTRTVLSCLPPDPADGRQPSAPAAPFAAAVTRSLTRGATALHTMGEILGWLTDER
                     SRRELVQQRVPLEETAFSGWRRDDHAIAHKDGDYFRVIGVSRASSREVSSWSQPLLA
```

CDS
```
              PVPGPGLAAFVTRRIRGVLHVLHARTEAGLLNGPEMAPTVQCRPLNYRAVPAEYRPAY
              LDVVLSADPGRIRYDTLQSEEGGRRFHHAENRYVVVEAEDDFPVEVPRDFRWLTLHQIL
              ALLHHSNYNVEARSLVACIQALS"
              2933..3541
              /note="Orf9; involved in the 3,5-epimerization step in
              dTDP-L-oleandrose biosynthesis"
              /codon_start=1
              /transl_table=11
              /product="dTDP-4-keto-6-deoxyglucose 3,5-epimerase"
              /protein_id="AAF59933.1"
              /db_xref="GI:7329193"
              /translation="MELLDVDGAWLYTPEIMRDERGEFLEWFRGRTFQEKIGHPLSLA
              QANCSVSRKGVLRGIHFADAPPGQAKYVTCASGTVLDVVDVVRRGSPTFGRWAAVRLD
              AARHQGLYLAEGLGHAFMALTDDATVVYLCSQPYVAEAERAVDPLDPAIGIEWPTDID
              IVLSAKDTHAPSLAQAAETGILPDYEECRAYIAEAAGVGPAR"
```

CDS
```
              3538..4608
              /note="Orf8"
              /codon_start=1
              /transl_table=11
              /product="dTDP-D-glucose synthase"
              /protein_id="AAF59934.1"
              /db_xref="GI:7329194"
              /translation="MKALVLAGGSGTRLRPITHTSAKQLVAVANKPVLFVGLEAIAAA
              GITDVGLIVGDTAGEVRAAVGDGAKFGLDITYIEQSRPLGLAHAVLIAHTYLGDDDFV
              MYLGDNFIVGGIDDLVRTPRDGRPPAARILLTHVSDPSAFGVAELDDDGRVVGLEEKP
              RHPKSDLALVGVYFTPAIHEAVRAIEPSWRGELEITHAIQHLIDNGADIQSMVIEGY
              WKDTGNVADMLEVNRTVLEDLEPRIEGTVDEHTVIGRVVVGEGARVTNSRIGMPAII
              GAGAEISDSYIGPFTSVGDNCRITGSEMEFSIMLAESAITGVRRIEGSLIGRNVQVTQ
              SLHAPNAHRFVLGDHSKVEIQS"
```

CDS
```
              4605..5612
              /note="Orf7"
              /codon_start=1
              /transl_table=11
              /product="dTDP-D-glucose 4,6-dehydratase"
              /protein_id="AAF59935.1"
              /db_xref="GI:7329195"
              /translation="MNLLVTGAAGFIGSRYVHHLLEAARAGGEPAPVITVLDKLTYAG
              VLGNVPDDPAVTFVRGDIADAPLVDSLMAEADQVVHFAAESHVDRSITSPGTFVRTNV
              LGTQMLLDACLRHGVGPFVHVSTDEVYGSIEHGSWPEHQPLCPNSPYSASKASSDLLA
              LSYHRTHGLDVRVTRCSNNYGPHQFPEKLVPLFVTNLLDGLRVPLYGDGLNVREWLHV
              DDHCLGVDLVRTQGRPGEVYHIGGGTELTNRDLTGLLLDAGVGVWDDHVPVADRKGHD
              RRYALDCAKAADELGYRPRRDFAEGIARTIDWYRDNRAWWEPLKKRPAPGPAARPGAAA
              P"
```

CDS
```
              5609..6493
              /note="Orf6; involved in the 4-ketoreduction step in
              dTDP-L-oleandrose biosynthesis"
              /codon_start=1
              /transl_table=11
              /product="4-ketoreductase"
              /protein_id="AAF59936.1"
              /db_xref="GI:7329196"
              /translation="MRWLITGAAGMLGRELVRRLAENEEDVAALGHDHLDVTRPFAVR
              AALAEHRPGIVVNCAAYTAVDDAETDEAAAALLNAEAPRLLAEACAPHGARLVHLSTD
              YVFPGDARTPYAEDHPTIADPRSAYGRTKRDGEQAVLTPTATVLRTAWLYGRTGRSFV
              RTMIEREARRGGAIDVVADOCGOPTWTGDLADRIIAVGRHPGVHGILHATNAGSATWYD
              LAQEVRLLDADPGRVRPTTGAAAFRRPAPRAYSVLGHDRWRGTGLAPLRDWRSALRE
              AFPDILAAEHPPTRRGAA"
```

CDS
```
              6490..7671
              /note="Orf5; possibly involved in the 4-deoxygenation step
              in dTDP-D-desosamine biosynthesis"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="3,4-dehydratase-like protein"
              /protein_id="AAF59937.1"
              /db_xref="GI:7329197"
              /translation="MKRGVHDLALFGGDAAFLQPLYMGRPNTGDRKRLLDRLEWALDN
              RWLTNGGPLVREFEQRIADLAGVRNCVATCNATAGLQLLLREAEVTGEVIMPSMTFVA
              TAHAVRWLGLRPVFCDIDPDTGCLDPKLVEAAVTPRTGAILGVHLWGRPSPVDELAAI
              AAEHGLKFYDAHALGCTSRQRRLGSFGDAEVPSFHATKVVNSFEGGGIVTDDDTPA
              DALRALHNFGLGHDVGAGINAKMSEAAAAMGLTSLEAFADAVASNRANYELYRQELS
              GLPGVRLSSYDPAERNYHYVIALIDAGVTGLHRDLLTLRAENVVAQPYFSPGCHQ
              REPYRTEHPVSLPHTEHLAEQVIALPTGPAVSREDIRRVCDIIRVAAAHGPRITAQAG"
```

CDS
```
              A"
              7792..9249
              /note="Orf4; possibly involved in the 4-deoxygenation step
              in dTDP-D-desosamine biosynthesis"
              /codon_start=1
              /transl_table=11
              /product="unknown"
              /protein_id="AAF59938.1"
              /db_xref="GI:7329198"
              /translation="MALQVDRTPLDPIAVCAWPGGNSAAAALRPLIEEDAPGTGLEP
              DRLAEHLIALARRYGTEPFTPLESARRGLGLDRATFARVLAVFHRTPALRTAVERRPA
              GMWTNTLPLERRGVLDAAVREPAFPYSVGLYPGPSCMFRCHFCVRTGARYQQSA
              LTDGNAMFASLIDRMPTDNPHALYLSGGLEPLTNPGTGDLVRRAAARGFKLSLYTNSF
              ALTRQTLDRQPGLWDLYALRTSLYGLSEDDYVATTKKGAFQRVKDNLTRFQALRRER
              EAPVRLGLNYIILPGRAGRLTGLADYFADLNDAAPDRPVDFLTLREDYSGRPDGKLAP
              EERVELEHGLAAFEERIRTRAPSLHVDYGYALQSLRLGVDAELPRIRPETMRPTAHPQ
              VAVQVDLLGDVYLYREAGFPGLQGAERYVAGRILTTGTLEEVVRRFVTEGRQVAPRPG
              EEYFLDGFDQTVTARLNQMETDIADGWAEHRGFLR"
```

```
BASE COUNT      1482 a      3627 c      3090 g      1324 t
ORIGIN
```

```
Query Match               14.4%;  Score 986.2;  DB 1;  Length 9523;
Best Local Similarity     67.9%;  Pred. No. 1.2e-66;
Matches 1410;  Conservative   0;  Mismatches 658;  Indels   9;  Gaps   2;
```

```
Qy   2262 GCCGAGGTGAAGGCCCTGCTCCTGGCAGTGGAACCGCAGCAGACTGAGGCCCTTCACC 2321
               |  |||||||||||| |||| ||||| || |||| || ||||| |||| |||||| |||
Db   3532 GGCCCGGTGAAGGCCCTCGTACTGGCCGGCCACCCGGCCACCGGGCTGCCGGCGATCACC 3591

Qy   2322 CACACCGCGCCAAGCAGCTGCTCCCATCGCCAACAAGCCCGTGCTCTCTCACGCGCTG 2381
               ||||||  |||  || |||||| ||||||| || ||||  || ||| ||| |||| |||
Db   3592 CACACCTCCGCCAAGCAACTGGTCGCCGTCGCGAACAAACCGGTGCTCGTGGGCGTACGGCCGG 3651

Qy   2382 GAGTCCCTGCCCGGCGGGTGTCCGGAGGCCGGCGCCCTGATCGGCCGCGCGGCG 2441
               |||||| | ||  || || ||||| || || ||||| ||| | ||||| | || ||||
Db   3652 GAGGCCATCGGGCCGCGGCCGGATCACCGGCGACGGACCGGCCGCGCGACCACGGCCGGC 3711

Qy   2442 GAGATCCCGGAACTCACCGCGCACGGACCACCCGTTCGGGTTACGGCATCACCTACCTCCAC 2501
               ||| || || |||| |||||  ||  || ||||| | ||||| |||||| || ||  ||
Db   3712 GAGGTCGGCGACGGCCAGTTCGGCCTCGACATCACCTACCATCGAG 3771

Qy   2502 CAGCCCCGCCGCTCGGTCGCCGCACGGGTGCCATCGCCCCGGCTTCCTGGGCGAC 2561
               ||  || || || ||||| || |||||| ||||||||| ||||||| |||| ||||
Db   3772 CAGAGCCGCCTCGGCGGCTCGGGCCTCGACCGCGACCACCTGCGACCTCGGGCGAC 3831

Qy   2562 GACGACTTCCTGCTGTACCTGGGGGACAACTACCACCTGCCCCAGGGCGTCACCGGACTTCGCC 2621
               ||||| || | || ||||| ||| |||||| ||||||| |||||| | | |||| || |||
Db   3832 GACGACTTCGTCATGTACCTGGGCGACAACTTCATCGTCGGCGGCATCGACGACGTGGTC 3891

Qy   2622 CGGCCAAT---CGGCCGGCCATCCGGC---CGGCCGTCGCCCGGCTGCGCGACCCGGTGCGCGGAC 2678
               ||| || |        | || ||  |   ||||| ||| ||||||||| || |||| |||
Db   3892 CGGACCTTCCGGCGACGGCCAGGCCGGCGGCGCCGGGATCGTCGGCTGCTCACCCATGTGTCCGAC 3951

Qy   2679 CCGTCCGGCCTTCGGCGTCGGCGGAGGTCGACGCGGACGGGAACGTGCTCGGCCTTGGAGGAG 2738
               || | ||  | |||||||   |  ||||||  ||||||| || |||||||||| || |||||
Db   3952 CCCAGCGGCCTTCGGCGTCCGGCGTCCGACGACGGGCCGGGGTCGTGGGCCTTGGAGGAG 4011

Qy   2739 AAACCCGACGTCCCGGCCAGTCGCTCGCTCGCGGTGTACGGCTTCAGCCCGGCC 2798
               || |||  ||| ||| || |||   |||||| ||| ||||| || ||||||| |||
Db   4012 AAGCCCCGGCCAGGACGCGACCTGGCACTGTCGGCGCGGTCTACTTCTTCACCCCCGCC 4071

Qy   2799 GTCCACGAGGCGGCTACGGGCCATCACCCCCTCCGGCGGGATCAGCACCCAC 2858
               ||||||||| | ||||  || |||||  |||  |||||| ||||||||||||||
Db   4072 ATCCACGAGGCCGTCCGCGCCATCGAACCCTCCTGGCGCGGGGAACTGGAGATCACCCAC 4131

Qy   2859 GCCGTGCAGTGGATGATCGACCGGGGCCTGCGCCGTACGGCGGACACCACCCGGCCC 2918
               || | ||  | ||  ||  || |||  | ||||| |  ||||| | |  |  ||| |||
Db   4132 GCCATCCAGCACCTCATCGACAACGGCGCCGACATCCAGTCCATGGTCATCGAGGGCTAC 4191

Qy   2919 TGGCGGCGACACCGGCAGCCGGAGGACATGCCGAGGATGCTGGAGGTCCTGGACGGA 2978
               |||  |||||  ||| ||||| || ||||| || |||  | |||| | |||||| |||
Db   4192 TGGAAGGACACGGGCAACGTCGCCGACATGCTCGAGGTCAACCGGACCGTGCTGGAGGAC 4251
```

```
Qy   2979  CTGGAGGGCCGCATCGAGGGAGGAAGGTCGACGCGCCACAGCACGCTGGTCGGCGGGTCCG  3038
                 ||||||  ||||||||| || || || |||||| | |||| ||||| ||||
Db   4252  CTGGAGCCCCGCCATCGAGGGACGAGCACGTGGACGAGCACCACCGCTGTGATCGGCAGGGTCGTG  4311

Db   5326  GCGGGCGGCCACCGAACTGACCAACCGGGACTTCACGGCGCCTCCTGCTTGTGACCGGTCCGGCG  5385

Qy   3039  GTGGCCGAAGGCGGCGATCGTGCGGGGTTCACACGTGGTGGGCCCCGGTCGTGTGATCGGCGCG  3098
                  ||| ||| |||||| ||| ||||   ||||| |||||| || |||||||
Qy   4119  CGGGCCCGGAGCGCATCGTCCACGTCGAGAACCGCAGAGACCGCAAGGGCACGACCGGCCTACCGCGG  4178

Db   5386  TGGGCTGGGACGTCGTGGGACCCCGGTCGCGGACCCGCCAAGGGCCAAGGGCCGCTACCCC  5445

Qy   4179  TCGACCACAGCAAGATCACCCGCGAACTCGTTACCGCCGCCGCCGACTTCGGCACCG  4238

Db   5446  TCGACTGCGCCAAGGCGCGGACGAGCTGGGCTACCGCCGCGGGACTTCGGCGGAGG  5505

Qy   4239  CGCTGCCGCGACACCGGCGAAGTGGTACGAGCCGCACGAGGACGGTGGCGCGTCCCCTGCTCG  4298

Db   5506  GCATCCGCCCGACCACCATCGACTGGTACCGCGACAACCGCCGCTGGTGGGAACCCCTGGAAGA  5565

Qy   4299  CCGGCGACATGACGTCGGCCGACGACCGCAACCACCGGC  4335

Db   5566  AGCGGCGCGGCCGTCCCGGTCCCCGCCGCCCCGAGGGGCAGC  5602
```

```
RESULT 4
SAMTMDE
LOCUS           SAMTMDE                  2404 bp    DNA         linear   BCT 02-MAR-1998
DEFINITION      S.argillaceus mtmD and mtmE genes.
ACCESSION       Y10907
VERSION         Y10907.1  GI:2924298
KEYWORDS        mtmD gene; mtmE gene; TDP-D-glucose synthase;
                TDP-D-glucose-4,6-dehydratase.
SOURCE          Streptomyces argillaceus
  ORGANISM      Streptomyces argillaceus
                Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
                Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE       1  (bases 1 to 2404)
  AUTHORS       Lombo,F., Siems,K., Brana,A.F., Mendez,C., Bindseil,K. and
                Salas,J.A.
  TITLE         Cloning and insertional inactivation of Streptomyces argillaceus
                genes involved in the earliest steps of biosynthesis of the sugar
                moieties of the antitumor polyketide mithramycin
  JOURNAL       J. Bacteriol. 179 (10), 3354-3357 (1997)
  MEDLINE       97294479
  PUBMED        9150235
REFERENCE       2  (bases 1 to 2403)
  AUTHORS       Lombo,F.
  TITLE         Direct Submission
  JOURNAL       Submitted (30-JAN-1997) F. Lombo, Universidad De Oviedo, Biologia
                Funcional, Julian Claveria S/N, 33006 Oviedo, SPAIN
REFERENCE       3  (bases 1 to 2404)
  AUTHORS       Lombo,F.
  TITLE         Direct Submission
  JOURNAL       Submitted (27-FEB-1997) F. Lombo, Universidad De Oviedo, Biologia
                Funcional, Julian Claveria S/N, 33006 Oviedo, SPAIN
COMMENT         On Mar 4, 1998 this sequence version replaced gi:1814003.
FEATURES             Location/Qualifiers
     source          1..2404
                     /organism="Streptomyces argillaceus"
                     /strain="ATCC 12956"
                     /db_xref="taxon:41951"
     gene            126..1193
                     /gene="mtmD"
     CDS             126..1193
                     /gene="mtmD"
                     /codon_start=1
                     /transl_table=11
                     /product="TDP-D-Glucose synthase"
                     /protein_id="CAA71846.1"
                     /db_xref="GI:1814004"
                     /db_xref="SPTREMBL:O08245"
                     /translation="MKALVLSGGSGTRLRPITHTSPKQLVPVANKPVLFYALEDIAQA
                     GITDVGIVVGETARETEDAVGDGSKFGLQVTYLPQEKPLGLAHAVLIARDYLGDDDFV
                     MYLGDNFVVGGIAGLVDEFRADRDPDARILLTPVDPTAFGIAELDGEGRVVRLEEKPT
                     HPKSDMAVIGVLYLFTPLVHEAVRAIEPSWRGELEITHAIQWLIDNRHDVRSTVIHGYW
                     KDTGNVMDMLEVNRIVLEGVEPCTEGPVDAESEVIGRVRIEAGARVSNSRIVGPAVIG
                     SGTVLSGSYIGPYTSVSEGCRIEDSEIEFSIVLRDSRVRGVRVRGSLIGRSVTVAPA
```

```
Qy   3279  CCGCAGGCTCACCGACTGGTGTGATCGGCGACCACAGCAAGGTGTATCTCACCCCATGACCA  3338
                 || | ||||| ||||| |||||||||||||||||||||||| ||||   ||
Db   4552  CCCAACGCCACCGCTTCACCGACCACAGCAAGGTGAGAGATCCAGTCAT-----  4606

Qy   3339  CGACCATCTCGTCACCGGCGGAGCGGGCTTCATTCGTCCGCCTACGTCCGGCCGGCCTCC  3398
                 |  | ||||| ||||||||   || ||||||| || ||||| || ||
Db   4607  -GAACCTCCGTCGTCACCGGTCGCCCGGATTCATCGGCCTCCCGTACGTGCACCACCTGC  4665

Qy   3399  TGTCGGCCGGGGGGCCCCGGGCGGCGTCGCGGGAGCCCCGGTGACCGTCCGACAAACTCACCTACGGCG  3458
                 ||   |||  | | ||||| |||| | ||||| | | | |||| |||||||| ||||||
Db   4666  TGGAAGCAGGCGGCGGGGCGGGGGGAGCCCCGGTGATCACGGTGCTGGACAAGCTCA  4725

Qy   3459  GCAGCCTCGCCCGCGCTCGACGCGTGACCATCCCCGCCTTCGTCCAGGGGCG  3518
                 || ||||| ||||||| || ||||| ||| |||||| |||||| ||| |||
Db   4726  CCTACGCCGGTCCTCGGCCAACGTCCCCGACGACCCCGTCACCTTCGTGCGGCGGGTG  4785

Qy   3519  ACGTGTGCGACACCGGCTCGTCGACACGCTGGCCGTCGGCCACGACGACATCGTGCACT  3578
                 | ||||||||| |||||||||  | ||||| |||| || |||||||| | |||||
Db   4786  ACATCGCCCACGCACCACTGCTGGTGACTCCCTGATGGCCCAGGTGTGCACT  4845

Qy   3579  TCGCGGCCGAGTGCACGTGCACGCTCCATCACCGACGACGCGGTGCCTTCACCCGCCACCA  3638
                 |||| ||||| ||||||||| |||| ||||||||  | |||| ||||| ||||||||
Db   4846  TCGCCGCCGGAGTCGCACGTGCACGCTGGACCTGGACCTCCATCACCTTCGTGCGACCACCA  4905

Qy   3639  ACGTGCTGGCCACCCAGGTCTCGCTGCTCGACGCGTCGCCACGGTGCGCACTTCG  3698
                 |||||||||| |||||| ||||||| |||| || || ||| || ||||||||
Db   4906  ACGTCCTCGGGCACCAGGTCTGCTGCGTCGACGGGGCCGTTCG  4965

Qy   3699  TGCACGTCTTCACCGGACGAGGTGTACGGCCTCCCTCCCCGCACGGGGGCCGCCGGAGAGCG  3758
                 ||||||| ||||| |||| || |||||| |||  |||||||||||||  ||| ||||
Db   4966  TGCACGTGTCCACCGACGGGGTCTACGGGCTCCATCGAGCACGGATCCTGGCCGGCACC  5025

Qy   3759  ACCCCCTGCTTCCGACCTGCGTTCCGACCTCGCGCGGTCGAAGGCGGCCGCCATGGGGCGC  3818
                 |  |||||| || ||| || ||| || ||| | ||| |||||||||| || |||||||
Db   5026  AGCCGGCTGTGCGCCACCTCGCCAACTCGCCTCCCCACCTGCTCCGCGC  5085

Qy   3819  TCGCCCCACCACCGCACCTTCGG  3878
                 |  ||||||||||||| || ||
Db   5086  TGTCCTACCACCGCACCCCGGCGC  5145

Qy   3879  GCCCCCACCAGCATCCGAGAAGCTCATACGCCGCTTCCTGACCAGCCTTCCTGTCCGGCG  3938
                 |||   ||| |||||| || |||| |  ||||| | |||| | |||||| |  ||||||
Db   5146  GGCCGACCAGTTCCCCTCGACGGCGGGCCCTTCAACCTCCCGACTCCGACGGCC  5205

Qy   3939  GCACCGTTCCCCTCCACGGCCGACGGGCACGGGGCCGCTCACCAACCTCCTGCACGTCGACGACC  3998
                 ||   ||| ||| ||| ||||  | ||||| ||| ||||||| |||| |||| ||||||||
Db   5206  TGCCGGGTCCCCTCTACGGCGCACGCACGCGCTCAACCCCACGACTCCGACGAAC  5265

Qy   3999  ACGTCAGGGCGTCGAACTCGTCGGCGGCGCGGGAGATCTACAACATCG  4058
                 || ||||  | ||| ||||| || || ||  || || |||| || |||
Db   5266  ACTGCCTGGCGTCGACGTTCCGCACCATCGTCTACCACATCG  5325

Qy   4059  GGGGCGGCCACCTCGCTGCCCAACCTGGAGCTCACGCACCTGGTTGCTCCGCACTGTGCGCGCG  4118
```

```
gene            PRKPAGTRLVLGDHSEGQISS"
                1190..2185
                /gene="mtmE"
CDS             1190..2185
                /gene="mtmE"
                /codon_start=1
                /transl_table=11
                /product="TDP-D-Glucose-4,6,-dehydratase"
                /protein_id="CAA71847.1"
                /db_xref="GI:1814005"
                /db_xref="SPTREMBL:O08246"
                /translation="MTTSLIVTGGAGFIGSHYVRTLLGPRGVPDVTVTVLDKLTYAG
                TLTNLAEVSDSDRFRFVRGDICDAPLVDDLLAVHDQVVHFAAESHVDRSLLGAADFVR
                TNVITGTLLDALRQGIERFVHISTDEVYGSIDAGSWPETAPVSPNSLYSAAKASSD
                LVALAYHRTHGLDVRVTRCSNNYGSHQFPEKVIPLFVTSLLIDGREVPLYGDGTNVRDW
                LHVDDHVRAIELVRTGGRAGEVYNIGGGTELSNKELTQLLLDACGAGWDRVRYVTDRK
                GHDRRYSVDCTKIRRELGYRPAREFGDALAETVAWYRHHRAWWEPLTRAYGAVAA"
BASE COUNT      372 a      833 c      837 g      362 t
ORIGIN

Query Match      14.3%;        Score 977.8;  DB 1;   Length 2404;
Best Local Similarity   68.0%;    Pred. No. 8.6e-66;
Matches 1379;   Conservative   0;   Mismatches 647;   Indels    3;   Gaps    1;
```

```
Qy   2269 TGAAGGCCCTGGTCCTTGGCGCAGGTGAAACCGGCAGAGACTGAGGCCGTTCACCCACACCG 2328
          ||||||||||||||||||| ||| || ||||| |||||||||| ||||||||||||||||
Db    127 TGAAGGCACTCGTTCGTTCGGCTGGCGGTGGCACCGGCACGCCTGGGGCACACCACCACAT  186

Qy   2329 CCGCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTCTTCACGCGGCTGGAGTCCC 2388
          | |||||||||||||| ||| |||| |||||||| ||||| |||||||||||||||||
Db    187 CGCCCAAGCAGCTGGTGCCGGTGGCGAACAAACCTGTTCTCTTCACGCGACCCTGGAGGACA  246

Qy   2389 TCGCGGGCGGGGTGTCCGGAGGCCGGCGTCGTGCTCGGGCGCCGTACGGCCGGGAGATCC 2448
          |||| || || |||||||||| | |  |||| || |||| |||||||||||||||||||
Db    247 TCGCGGAGGCAGGGATCACCGACGTCGGGGATCGTCGGTCGTCGGTGAGAGCGCCCGAGATCG  306

Qy   2449 GCGAACTCACCGGCGACGGGCACCGCGCTTCGGTTACGCATCACCTCACCTCCACCAGCCCC 2508
          |||| || | | ||||| | |  |||| ||||||||||||||||||| |||||||| ||
Db    307 AGGACGCGGTGGGCGACGGTGGCCAAGTTCGGCCTCCAGGTGACCTACCTGCCGCAGGAGA  366

Qy   2509 GCCCGCTCGGTCTCGCGCACGCGGTGCCATCGCCGCCGGCTTCCGGGCGACGCAGCGACT 2568
          |  |  |||| ||  |||||||||||| ||| |||| ||||| || ||||||||||||||
Db    367 AACCCCTGGACTCGCCCACGCGTCCTGATCGCCGACTTCGTGACGACGACT  426

Qy   2569 TCCTGCTGTACCTGGGGACAACTACCTCGCCCAGGCGTCACCGACTTCGCCCGCCAAT 2628
          |||| ||||||||||| | |||||| |||||| ||| |  ||||||||||||||||||
Db    427 TCGTCATGTACCTGGAGACAACTTCGTCGTCGGTCGGCATCGCGCGTCGGTGGACCAGT  486

Qy   2629 CGGGCCGATCCCCGGCGCGGCTGCGTGTGCTCGGGCGTCGCGGACCCGTCGCGCCT 2688
          |||||| |  |||||||| ||  || ||| |||||| ||||||||||||||  |||
Db    487 TCCGTGCGGACCGGCCGACGCGGGCGATCCGGGCTTCCCGAGTCGGTGCTCCGACGGCCT  546

Qy   2689 TCGGGTCGCGGGAGGTCGACGCGGACGGAACGTGCTGCGCTTGGAGGAGAAAACCCACG 2748
          ||||  | | |||| ||||||||||||| ||  ||||||||||||||||||| ||||
Db    547 TCGGGATCATCGCGAGCTGGACGGGGAGGGCGGGTGGCGCTCTATTCCGCGGAGGAAGCCGACGC  606

Qy   2749 TCCCGCGCAGCTCGCTCGCGGCCTCATCGGGCGTACGGCGTGTACGCCGTCAGCCGCG 2808
          |||||||||||| ||||||||||||||| |||||||||||||||||||| ||| ||||
Db    607 ACCCCAAGAGCGACACATGCGGTGATCGGCCGTGATCGGCCTACCGTGTTCCACAGGCCGCCAGG  666

Qy   2809 CGGTACGGCCGATCCCCGGCGCTGCTCACCCCGGCTGGAGATCACCCACGGCCGTGCAGT 2868
          | || || ||| ||||||||| ||  || || |||||||||||||||| |||||||||
Db    667 CCGTGCGGCCGCATCGACGCCGTGGGCGGCGGAGTCATCCCGCGAAATCACCCATGGCCCATCCAGT  726

Qy   2869 GGATGATGCAGCGGGGCCTGAGTCGTACGGGCCGAGACCCCGGCGGCCCTCGGCCGGACA 2928
          ||||| |||| | | || || |||| |||||||||||||||||| ||| || |||||||
Db    727 GGCTCATCGACAACCCCAGACGTGCGGGTCCTTACCGTGATCCAGCGGCTACTGGACGCCGACTGA  786

Qy   2929 CCGGCAGCGCGGGAGGACATGCTGGAGTCAACCGTCACGTCCTGGACGACTGTGAGGGCC 2988
          ||||| |||| ||||||||| || ||||| ||||||||||||||||||||| |   |||
Db    787 CCGGCAGCGTCATGATGACGACGTCAACCGCCTACCGCGGCGTCGCGGAAGGTCTCACAGCTCGAGCCGGT  846

Qy   2989 GCATCGAGGTGGGGAAGGTCGACGCGCAGCACGCTGTGTCGCGCGGCCGAAG 3048
          |||| ||||||| ||  |||  | |||||||||||||||||||||| |||
Db    847 GCACCGAGGGCCCGGTCGCCGGAGAGCGGAGTGATCGATGCGGGCCGGCGGCCG  906

Qy   3049 GCGCGATCGTGCGGGGTCACACGTGGTGGGCCCGGTCGGCCGGGGTGCCGGTCG 3108
          |||| || | | |  ||| || || ||||| ||| |||| ||||| |||||||
Db    907 GCGCCGGTCAGCAATTCGGCGATCGGCGGGGTGATCGGGTGATCGGGGTCGGCACGGTGA  966

Qy   3109 TCAGGCAACTCCAGTGTCCGCACCTCACTCGGCGGAGGACTGCCCGGGTCGAGGACA 3168
          ||  ||| | |||| ||| || |||| || | ||| |||||||||||||||||||
Db    967 TCAGCGGTTCCTACATCGGACGTCCGCTACACGCTCGACTAGGCCCTGGACCCT 1026

Qy   3169 GCGCCATCGAGTACTCCGTCTGCTGCGCGGACCGGCGTCGAGGGGGGCGTCCCGCATCG 3228
          ||| | ||||||| |||||| ||||| |||| ||||||||||| |||||| |||||||
Db   1027 GCGAGATCGAGTTCTGCGATCGTCTGCGACTCGTCGATCGTGCGGGGGGTCCCAGCGGGTCA 1086

Qy   3229 AGGCGTCCTCATCGGCCGCCGGCGCCCGTCTCGGCCGCGCGCCCCCCGTCTCCCCGCAGGCTC 3288
          |||||| || |||||||||||| || |||||||||| ||||||||||||||| ||||||||
Db   1087 GGGGTCCCTCATCGGCCCGCAGTGTCACCGTGGTCCCCCCCCGCAAGCCGGGCCGGCA 1146

Qy   3289 ACCGACTGTGATCGGCGACCACGCAAGGTGTATCCGCCGCCCATGACCACCA---CCAT 3345
          ||| | | | | | || || ||||||||| ||||||| ||| ||| || ||   ||||
Db   1147 CCCGGCTCGTTCTCCGGACGCAGCACCAGCGAGATCTCCTCATGACCACCACCATCCAT 1206

Qy   3346 CCTGTCACCGGCGGAGCGGCTTCATTCGTCCCGCCTACGTCGCCGGCCTCCTCGCGC 3405
          ||| |||||| ||| ||||| |||||| |||| ||| |||||||||||||||||||
Db   1207 CCTGGTGACGGGGGGCGGCGGCTTCATCGTCCGGCTCGCCACTACGTGCGGACCCTGCCGGTCC 1266

Qy   3406 CGGGGCCCCCGGGCGCGTCGCGGTGACCGTGCTCGACAAACTCACCTACGCCGCCAGCCT 3465
          ||||| || | ||||||||| |||| || ||||| |||| ||||||| |||||||||
Db   1267 GCGGGGGCGTGCCGGACGTGACGGTCTCGACAAGCTGCTCGACTACGCCGGGACCCT 1326

Qy   3466 CGGCCCCTGCAGGCGGTCGTGCTGTCGACCATCCGGCGGCGGCGGAGATCCTGTCGTG 3525
          ||||| | || | || |||| |||||| ||||| |||| ||||||  |||| |||||
Db   1327 CACCAACCTCGCCGAGGTGAGCGACAGCGACCGGTTCCGGTTCGTCCGGGGCGACATCTG 1386

Qy   3526 CGACACCCGGCTCGTCGACACGCGCTGGCGCGACTCACCTCCACCAGCCCC 3585
          ||||| || || ||| ||||||| |||||||| | |||| |||||||| |
Db   1387 CGACGCACCCCGGTCGACGACCTGCTCGCCGTGCACGACCAAGTGGTCCACTTCGCCGC  1446

Qy   3586 CGAGTCGACGTCGACCCGCTCCATCACCGACACAGCGGTGCCTCCCCACCAACGTGCT 3645
          |||| |||||||||| ||| ||| ||| || |||||||||| |||||  ||| ||||
Db   1447 CGAGTCGCACGTCGACCGGTCCCTGCTCGGCGGCCGCCGACTTCGTGCGACCAATGTCAC 1506

Qy   3646 GGGCACCCAGGTCGACTGCCTGGGGACAACTACCTCGCCCACGGCGTGCACTTCGTCACGT 3705
          ||| ||| |||| || ||||| || | |||||| |||| ||| |||| |||||||||
Db   1507 CGGTACCGGACAGCGCTCCCTCGACGCCTGGGATCGGAGTCGGGGACGTTCGTCACACAT 1566

Qy   3706 CTCCACCGACGAGGTGTACGGCGCCGCCGGAGAGCGGCGACCCCT 3765
          |||||| || ||| | ||| ||| ||||||  |||| ||||||
Db   1567 CTCCACCGACGAGGTGTACGGCTCCATCGACGCGGGCTCGTGGCCGGTCAGT 1626

Qy   3766 GCTTCCGACCTCGCCGTCCATGGCGCGTCGGCCGTCATGGCGCGTCGCCCA 3825
          |||| |  |||| ||| |||| ||| |||||||| |||| | || ||||
Db   1627 CAGCCCAACTCGTCGTCTATTCCGCGGCGGAGGAGCCGACGCCGTCGGCCTA 1686

Qy   3826 CCACCGCACCCAGGCTGGGGCCGAGACCCTGTCGCCGACTTCGGGTCGCCCCCA 3885
          |||||||||||||||  ||| || | ||  |||| || | |||||| |||||
Db   1687 CCACCGCACCCAGGGCGCGTGCTGCAGCCGGACTACCAGGGTCCCA 1746

Qy   3886 CCAGCATCCCGGAGAGCTCATACGGCTGGACGCCTCCTGCTCCGTCGCCACCGT 3945
          |||| ||| | ||||| ||| ||| ||||| | || |||||||||  ||||||
Db   1747 TCAGTTCCCGGAGAAGGTCATCCCGTGATCGCAGCCAGTCGCACCAGTCCATCCAGT 1806

Qy   3946 TCCCCTTACGGCGACGGGGACTGGAGGCAGCGCGTGCACGTCGACGACGACA 4005
          ||||||| ||||| |||| | | || |||| ||||| |||||||||||||
Db   1807 GCCGTTGTACGGGACGGACCACCAACGTCCGGGATCGGGCGGGGGCCGACGAGCTCCG 1866

Qy   4006 GGCGTCGAACTCGTCCGGTGTCGGTGGGGGCCGGGAGAGATCTACAACATCGGGGGCGCG 4065
          ||||| |||||  ||  |   || || | |||| ||  |||| || ||||||||||||
Db   1867 CGGATCAGCGTCGTCACGACCCAGCGGTCACCGGACGTCGGCGTACAACATCGGCGGCGGG  1926

Qy   4066 CACCTCGCTGCCCAACCTGACGCGTGCTCACTGTGCGGCGGTCGGTCCGGTGGCCGAAG 4125
          ||||  | | | |||  ||||
Db   1927 ...
```

```
Db    1927  CACCGAGCTGTCCAACAAGGAGCTGACCGCAGCTGCTCCTGGACGCCCTGTGGGCCGGCTG 1986
                   ||||||||||||||||||||| |||||| ||||||||||| | |||||
Qy    4126  GGAGGCCATCGTCCACGTCGAGAACCGCAAGGGCACGAACCCGCCTACGCCGGTCGACCA 4185

Db    1987  GGACCGCCGTGCGGTACGTCACCGACCCCAAGGGCACGACCCGGCCGCTACTCCGTCGACTG 2046
                   ||||||||||||||||||||||||||||||| | |||||
Qy    4186  CAGCAAGATCACCCGGAACTCGGTTACCGGCCGCGCACCGACTTCGCGACCCGCCTGGC 4245

Db    2047  CACCAAGATCCGGCCGTGAGCTCGGGTACCGGCCGCCGGCTCGGGGAGTTCGGGACCCACTGGC 2106
             || | ||||| |||||||| |||||||| | |||||| | |||||||||||
Qy    4246  CGACACCCGCAAGTGGTACGAGGCGCACGAGGACTGGTGGCGTCCCCTG 4294

Db    2107  CGAGACTGTCGCGCTGGTACCGCCACCACCGTGCCTGGTGGTGGGGAGCCGCTG 2155
```

```
RESULT 5
SAR7932
LOCUS       SAR7932               18877 bp    DNA     linear   BCT 26-FEB-2002
DEFINITION  Streptomyces argillaceus mithramycin biosynthetic genes.
ACCESSION   AJ007932
VERSION     AJ007932.2  GI:8926774
KEYWORDS    acyl carrier protein; acyl CoA ligase; aromatase; cyclase;
            D-mycarose 2,3-dehydratase; D-mycarose 3-C-methyltransferase;
            D-olivose 2,3-dehydratase; D-olivose 4-ketoreductase; D-olivose
            2,3-dehydratase; dTDP-glucose 4,6-dehydratase; dTDP-glucose
            synthase; ketoacyl synthase; ketoreductase; mtmC gene; mtmD gene;
            mtmE gene; mtmK gene; mtmL gene; mtmOI gene; mtmOII gene; mtmOIII
            gene; mtmP gene; mtmQ gene; mtmS gene; mtmTI gene; mtmTII gene;
            mtmTIII gene; mtmU gene; mtmV gene; mtmX gene; mtmY gene;
            oxygenase.
SOURCE      Streptomyces argillaceus.
  ORGANISM  Streptomyces argillaceus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Lombo,F., Blanco,G., Fernandez,E., Mendez,C. and Salas,J.A.
  TITLE     Characterization of Streptomyces argillaceus genes encoding a
            polyketide synthase involved in the biosynthesis of the antitumor
            mithramycin
  JOURNAL   Gene 172 (1), 87-91 (1996)
  MEDLINE   96257259
   PUBMED   8654997
REFERENCE   2
  AUTHORS   Lombo,F., Siems,K., Brana,A.F., Mendez,C., Bindseil,K. and
            Salas,J.A.
  TITLE     Cloning and insertional inactivation of Streptomyces argillaceus
            genes involved in the earliest steps of biosynthesis of the sugar
            moieties of the antitumor polyketide mithramycin
  JOURNAL   J. Bacteriol. 179 (10), 3354-3357 (1997)
  MEDLINE   97294479
   PUBMED   9150235
REFERENCE   3
  AUTHORS   Prado,L., Lombo,F., Brana,A.F., Mendez,C., Rohr,J. and Salas,J.A.
  TITLE     Analysis of two chromosomal regions adjacent to genes for a type II
            polyketide synthase involved in the biosynthesis of the antitumor
            polyketide mithramycin in Streptomyces argillaceus
  JOURNAL   Mol. Gen. Genet. 261 (2), 216-225 (1999)
  MEDLINE   99200478
   PUBMED   10102355
REFERENCE   4
  AUTHORS   Gonzalez,A., Remsing,L.L., Lombo,F., Fernandez,M.J., Prado,L.,
            Brana,A.F., Kunzel,E., Rohr,J., Mendez,C. and Salas,J.A.
  TITLE     The mtmVUC genes of the mithramycin gene cluster in Streptomyces
            argillaceus are involved in the biosynthesis of the sugar moieties
  JOURNAL   Mol. Gen. Genet. 264 (6), 827-835 (2001)
  MEDLINE   21148776
   PUBMED   11254130
REFERENCE   5
  AUTHORS   Salas,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1998) Salas J.A., Biologia Funcional Universidad
            de Oviedo, Julian Claveria S/N, Asturias, SPAIN
  REMARK    replaced by [6]
```

```
REFERENCE   6  (bases 1 to 18977)
  AUTHORS   Salas,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2000) Salas J.A., Biologia Funcional Universidad
            de Oviedo, Julian Claveria S/N, Asturias, SPAIN
COMMENT     On Jul 5, 2000 this sequence version replaced gi:3334818.
FEATURES             Location/Qualifiers
     source          1..18977
                     /organism="Streptomyces argillaceus"
                     /strain="ATCC 12956"
                     /db_xref="taxon:41951"
     gene            126..1193
                     /gene="mtmD"
     CDS             126..1193
                     /gene="mtmD"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="dTDP-glucose synthase"
                     /protein_id="CAA07754.1"
                     /db_xref="GI:3334819"
                     /db_xref="SPTREMBL:O08245"
                     /translation="MKAIVLSGGSGTRLRPITHTSPKQLVPVANKPVLFYALEDIAQA
                     GITDVGIVVGETAREIEDAVGDGSKFGLQVTYLPQEKPLGLAHAVLIARDYLGDDDFV
                     MYLGDNFVVGGIAGLVDEFRADRPDARILLTPVPDPTARGIAELDGEGRVRLEEKPT
                     HPKSDMAVIGVYLFTPLVHEAVRAIEPSWRGELEITHAIQWLIDNRHDVRSTVIHGYW
                     KDTGNVMDMLEVNRTVLEGVEPCTEGPVDAESEVIGRVRIEAGARVSNSRIVGPAVIG
                     SGTVISGSYIGPYTSVSEGCRIEDSEIEFSIVLRDSRVRGVQRVRGSLIGRSVTVAPA
                     PRKPAGTRLVLGDHSEGQISS"
     gene            1190..2185
                     /gene="mtmE"
     CDS             1190..2185
                     /gene="mtmE"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-glucose 4,6-dehydratase"
                     /protein_id="CAA07755.1"
                     /db_xref="GI:3334820"
                     /db_xref="SPTREMBL:O08246"
                     /translation="MTTSILVTGGAGFIGSHYVRTLLGPRGVPDVTVTVLDKLTYAG
                     TLTNLAEVSDSDRFRFVRGDICDAPLVDDLLAVHDQVVHFAAESHVDRSILGAADFVH
                     TNVTGTQTLLDAALRQGIETFVHISTDEVYGSIDAGSWPETAPVSPNSLYSAAKASSD
                     LVALAYHRTHGLDVRVTRCSNNYGHQPEKVIPLFVTSLLDGREVPLYGDGTNVRDW
                     LHVDDHVRAIELVRTGRAGEVYNIGGGTELSNKELTQLLLDACGAGWDRVRYYTDRK
                     GHDRRYSVDCTKIRRELGYRPAREFGDALAETVAWYRHHRAWWEPLTRAYGAVAA"
     gene            complement(2604..3365)
                     /gene="mtmTII"
     CDS             complement(2604..3365)
                     /gene="mtmTII"
                     /codon_start=1
                     /transl_table=11
                     /product="ketoreductase"
                     /protein_id="CAA07756.1"
                     /db_xref="GI:3334821"
                     /db_xref="SPTREMBL:O86480"
                     /translation="MDMGLTGRRVLVTGSSGIGAAVARAYADEGARVALTYRGGDEA
                     AERLARELGAGRHAAMAVRYSLGEPDTVESAVATVTERWGGVDVLVANAVWRWRAP
                     GTRFEEVASGDWRGFLDDNLAPTLRTVQLVVAGMRARSWGRIVLISSHVALDGHRGQE
                     FYGAAKSALHGFARSLAWDVGRDGVLVNVVCPGLTTTERVLTGLPDEIRERELGSTPT
                     GRLSSPEDIAGAVLFLGSAANANTTGQTLIVSGGR"
     gene            complement(3382..4977)
                     /gene="mtmOII"
     CDS             complement(3382..4977)
                     /gene="mtmOII"
                     /codon_start=1
                     /transl_table=11
                     /product="oxygenase"
                     /protein_id="CAA07757.1"
                     /db_xref="GI:3334822"
                     /db_xref="SPTREMBL:O86481"
                     /translation="MADPTRVLVAGAGPVGLTAAHELARRGLRVRLIDAAPRPAATSR
                     AVATHPRTLETYDQMGVIEILENAQQIRAFTMFSRGRRLVRLDADYTMPTRFPFTA
                     IVEQVDTEAVLRSAVARLGVDVEWETRLTGFSQDAEGVDVTLEHADGTTESTRVPWLV
```

```
                                                                    GCDGGHSTVRKLLGLPLLGESSETWMLADAPVDVDLPRDSIYWVHTGSQALMMVPFRR
                                                                    GGRWRLLDTTTAEPGTDAAVIADRFTRNSAPASAVRQVRTPTWTSVFTFQQRMVPRMG
                                                                    EGRVPVAGDAAHVHSPASRGMGMNTGVQEAYNLAWKLALVAEGHAERELLDSYSLERVP
                                                                    IGERILLGSTKKATFLVQLKNAFAPIALPVVFGLIRNVPPLRRKVQRQVLGGISGLQLD
                                                                    YPDSPLTRPAATGAGRPGPTRASVSGGADPADPGCRAWAEELRDTRWTLAVTPDAVAG
                                                                    APAANSWLGSVRTLGDGSQGPGPLADPNGRLRAALGLASGGWLLVRPDGYVAARGRALG
                                                                    GADLRQALAAAGLRDEQHVLPHS"
                   gene      complement(4970..5269)
                             /gene="mtmOIII"
                   CDS       complement(4970..5269)
                             /gene="mtmOIII"
                             /codon_start=1
                             /transl_table=11
                             /product="oxygenase"
                             /protein_id="CAA07758.1"
                             /db_xref="GI:3334823"
                             /db_xref="SPTREMBL:O86482"
                             /translation="MVTFVNKLTVNGDIEEFLTIKRSLTEFMSAQPGYRSSTLRQVG
                                                                    RPDVFIELAEWDDAASHQAAVRSEEFQARVKGLGALASVDPGLYEEASEDGAGRG"
                   gene      complement(5313..6857)
                             /gene="mtmL"
                   CDS       complement(5313..6857)
                             /gene="mtmL"
                             /codon_start=1
                             /transl_table=11
                             /product="acyl CoA ligase"
                             /protein_id="CAA07759.1"
                             /db_xref="GI:3334824"
                             /db_xref="SPTREMBL:O86483"
                             /translation="MSELPPGPAVPLDGLLRLAAERHPEQVAVRDERGAVDYAGLDAR
                                                                    ADRFARALLRLTGGRPTAVGVASVLDPVFAAAFYGTSRSGNRVVLVNPLYVREPVLEHV
                                                                    FRTAGIEIALVSAETRSPGGRRAALPDLREYVVDADRTGPPPGTRPLDDLLTEDGT
                                                                    AALPDPAGVDLDSVVCVQFTSGTTGPPKGVRLTHRNLVANAAQRAHALGLDAGSVCLN
                                                                    HLPLYHVMHLDSAVTAGATQVLCHDPDPVASVAAAEAGATHYFGLPVRLARLAADPR
                                                                    LASVVPGPELGLVRSGGSRLAPAVAARLRERLGVPVIQGYGLAELSPLSHNDRPERSK
                                                                    PGSVGPAVPGTECRIVDLETGAALDPGRPGEVLLRGPQLMAGYLGLPDAPDIDAAGWF
                                                                    HTGDVGYQDEDGWLFLVDRIKDVFKVDNELLVSPSEIEQVLLQDPDVADCVVADLPDEF
                                                                    SGAVVWAGVVPGVDLNPIVARANALLSDHQRIRRAERLTAVPRSPNGKTERTRL
                                                                    RERLRA"
                   gene      complement(6850..7797)
                             /gene="mtmQ"
                   CDS       complement(6850..7797)
                             /gene="mtmQ"
                             /codon_start=1
                             /transl_table=11
                             /product="aromatase"
                             /protein_id="CAA07760.1"
                             /db_xref="GI:3334825"
                             /db_xref="SPTREMBL:Q53669"
                             /translation="MPNSQQYSVRDEILVAAPAERVRSLVDDVVGWTQLHGPAVHAHV
                                                                    LRDTGDERLIRHWAVTGRNSVRTWTARHTRSADASRISFAHVMPEAPFASLEGAWTFR
                                                                    ERDGKTRVEMRHDFTLSDGSAGPADELTARLREGSRAYLSTLGHAAENHEELDRLLIS
                                                                    FTDSLFIAGDVKDAYDLWEADKWEERIPHIAKLDLEEPAPGVQFFDMLTSTPDGSSH
                                                                    RTRSRVRTVBPDIIYKQTELPALLDGHTGHWKFTRTPEGVIAEARHTATVKPSALHI
                                                                    LGEGTTVQDARNYLRRVLSANSVSNLRLAKRAYAEERAGV"
                   gene      8009..8461
                             /gene="mtmX"
                   CDS       8009..8461
                             /gene="mtmX"
                             /codon_start=1
                             /transl_table=11
                             /product="cyclase"
                             /protein_id="CAA07761.1"
                             /db_xref="GI:3334826"
                             /db_xref="SPTREMBL:Q53670"
                             /translation="MTAPTETQPHSSAGFGSLYAEIQQFYARHMQLLDGAADEWAAD
                                                                    FTEDASFSVPTVPAPVVGRDALAAAVRRTAESLAST'GEQHRHWPGVFDVPRRDGSVV
                                                                    VHSYTTVYASPRGGESRVHRVCTCTDVLVREDGRLLVRSREVIRDDLR"
                   gene      8642..9910
                             /gene="mtmP"
                   CDS       8642..9910
```

```
Best Local Similarity   68.0%;    Pred. No. 4e-66;
Matches 1379;   Conservative   0;   Mismatches 647;   Indels   3;   Gaps   1;

Qy   2269  TGAAGGCCCTGGTTCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCGTCACCCACACCG  2328
                  ||||| ||| ||  ||| |  ||| || || | | |||  ||||||| ||||||||
Db    127  TGAAGGCACTCGTTCTGTCCGGTGCTCCGGCACACGCCCATCACCCACACAT  186

Qy   2329  CCGCCAAGCAGCTGCTCCCCATCGCCAACAGACCCGTCCTGCTCTCTCTACGGCGTGAGTCCC  2388
                  | ||||||||| | | |  |    |  | | | ||| |||||  ||||||
Db    187  CGCCCAAGCAGCTGGTGCCGGTGGCGAACAAACCTGTTCTCTTCTACGGCGCCTGGAGGACA  246

Qy   2389  TCGCCGGCGGCGGTGTCCGGAGGCCGGCCTCGTCGTGGGCGCGTACGGCGGGGAGATCC  2448
                  ||| ||| || |||||  |   || |||||| |||  || || ||| | || ||||||
Db    247  TCGCGCAGGCCGGGATCACCGACCTCGGGATCGTCGTCGGCGTCGTGAGCAGCGCCCGCGAGATCG  306

Qy   2449  GCGAACTCACCGGCGACGCACCGCGTTCGGGTTACGGCTCACCTACCTCCACCAGCCCC  2508
                  |  |  ||  |  ||| || |  ||||  | || | ||  | |  |  ||||| |||||
Db    307  AGGACGCGGTGGGCGGACGGTCCAAGTCGGCCTCCAGGTGACCTACCTCGCCCAGGAGA  366

Qy   2509  GCCCGCTCGGTCTCGCGCACGCGGTCGCGCATCGCCCGCCCGGCGCTTCCTGGGCGACGACGACT  2568
                  |||||| ||| || |||| ||| ||| | ||||| ||||||
Db    367  AACCCCTGGGACTCGCCCACGCGGTCCTGATCGGCCGACTACCTCGGTGACGACGACT  426

Qy   2569  TCCTGCTGTACCTGGGGACAACTACCTGCCCAGGCCGTCACCGACTTCGCCCGCCCAAT  2628
                  ||  | |  |||||||||| | | |   |  |   ||||| ||  ||  || |  | |||
Db    427  TCGTCATGTACCTGGGAGACAACTTCGTGTCGGCGGCCATCGACGTCGTGGTGACGAGT  486

Qy   2629  CGGCCGCCGATCCCGCGGCGGCCCGACGCGCGGGATCCTGCTCACCCCCGGTCGCGACCCGTCCGCCT  2688
                  |  |  | |  | | ||| ||| |  |  ||| |   || || |||||||  | || |||| |||| |
Db    487  TCCGTGCGGACCGGCGGCCCGACGCCGTTCCGACCCCGTTCCCGATCATCGACGCGGCCCT  546

Qy   2689  TCGCCGTCCGGAGGTCGACGCGGAACGTGCTGCCGCTTGGAGGAGAAACCCGACG  2748
                  ||||| ||| || || | | |    |  ||| |  | | |||  ||||| |
Db    547  TCGGCATGCGACGCTGACGCGGAGGGCCGGGTCGTACGCCTGACGAAGCCCGACGC  606

Qy   2749  TCCCCGCAGCTCGCTCGCCTCATCGGCCGTACGCCTTCAGCCCGGCGTCCACGAGG  2808
                  ||  |||   | |   | ||| | |||||||||  |||| | | |||  ||||||| |
Db    607  ACCCCAAGAGCGACATGGCGGTGATCGGGCTCTACCTGTTCACCCCCGCTCGTGCACGAGG  666

Qy   2809  CGGTACGGCCCATCACCCCCTCCGGCCCGGCGAGCTCGGCAGGTCCAGT  2868
                  | || ||||||   ||||||    | |||| || |  | |||| ||| ||
Db    667  CCGTGCGGCCCATCGGCGTCGTCGCGGCGGGGACTGGAAATCACCCCATCCAGT  726

Qy   2869  GGATGATCGACGCCGGGCCTGCGCGTACGGGCCACCACCCGGCGCCGACA  2928
                  || | |    |  |  || |||  |||   ||| | |||| | ||||||||
Db    727  GGCTCATCGACAACCGCACGACCTGCGGTCTACCGTGATCACGGCTACTGGAAGGACA  786

Qy   2929  CCGGCAGCGGCGAGGACATCGTCAACCGTCAACCTGAGGTCAACCGCGAGGGGCC  2988
                  ||||| ||   || | | ||    |||||||||||| |  || ||  |  |  || ||
Db    787  CCGGCAACTCATCGACATGTCATGGACATGTCCGGAGGTCAACCGACACGGTGCTCGAGCCGT  846

Qy   2989  GCATCGAGGGGAAGTCGACGCGCACAGCACGTGGTCCGCCGGGTGCCGCGAAG  3048
                  ||  | || | || | |  ||| |   | | |  |||||| | |||| |  | |
Db    847  GCACCGAGGGCCCCGAGCACCGGAGAGCCAGGTGATCGGCGCCGCGTCAGGATCGAGGCGCCG  906

Qy   3049  GCGGGATCGTGCGGGGGTCACACCTGGTGATCGGCGCGGGTGCCCGTCG  3108
                  |||| |    |||| || |   |||| | | ||||| | |||||| |||| |
Db    907  GCGCCCGGGTCAGCAATCGTCGGATCGTGGTCCCCGGGTGATCGGGTGCGGGCACGGTGA  966

Qy   3109  TCAGCAACTCCAGTGTCGGCGGCCGTACACCTCCATCGGGAGGACTGCCGGGGTCGAGGACA  3168
                  ||||| || |   |||  |  | ||  | ||  |||  ||||  |  | ||||| |||||||
Db    967  TCAGCGGTTCCTACACGGGCCGTACACGTCCGTCCGAGGGGTGCCCGAGGGGCTGCCGAGGACA  1026

Qy   3169  GCCCATCGAGTACTCCGTCCTGCTCGTGCGCGGGCCCATCGGCGTCCCGCATCG  3228
                  |   | || | | |  | ||||  | |||||  ||||||| | | | |  |||||
Db    1027  GCGAGATCGAGTTCTCGATCGTGCTGCGCGGCCACTCACGGGTGCCGGGGGTCTCCAGCGGT CA  1086

Qy   3229  AGGGCTCCCTCATCGGCCGGCCCGCGTCGCCGGCGCCCCCGTCTCCCCGCCAGGGCTC  3288
                  | |||| ||   | |  |||| ||| |    |  |   ||| ||  ||| ||| |
Db    1087  GGGGGTCCCCTCATCGGCCCGGCCCCCGCAGTGTCACCCGTCGGCCCCCCGCAAGCCGGCCGGCA  1146

Qy   3289  ACCGACTGGTGATCCGCGACCAGCAAGGTGTATCTCACCCATGACCACGA---CCAT  3345
                  ||| || || ||  |||  ||    |  ||||||||| ||||||| ||   |  ||||
```

```
Query Match   14.3%;   Score 977.8;   DB 1;   Length 18977;
```

```
            4-ketoreductase (oleU), dehydratase (oleI), reductase (oleNI),
            glycosyltransferase OleI (oleI), aminotransferase (oleN2), and
            glycosidase OleR (oleR) genes, complete cds.
ACCESSION   AF055579
VERSION     AF055579.2  GI:5902165
KEYWORDS    .
SOURCE      Streptomyces antibioticus
ORGANISM    Streptomyces antibioticus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 9806 to 15052)
AUTHORS     Quiros,L.M., Aguirrezabalaga,I., Olano,C., Mendez,C. and Salas,J.A.
TITLE       Two glycosyltransferases and a glycosidase are involved in
            oleandomycin modification during its biosynthesis by Streptomyces
            antibioticus
JOURNAL     Mol. Microbiol. 28 (6), 1177-1185 (1998)
MEDLINE     9834380
PUBMED      9680207
REFERENCE   2  (bases 1 to 9805)
AUTHORS     Aguirrezabalaga,I., Olano,C., Allende,N., Rodriguez,L., Brana,A.F.,
            Mendez,C. and Salas,J.A.
TITLE       Identification and expression of genes involved in biosynthesis of
            L-oleandrose and its intermediate L-olivose in the oleandomycin
            producer Streptomyces antibioticus
JOURNAL     Antimicrob. Agents Chemother. 44 (5), 1266-1275 (2000)
MEDLINE     20233594
PUBMED      10770761
REFERENCE   3  (bases 9806 to 15052)
AUTHORS     Quiros,L.M., Aguirrezabalaga,I., Olano,C., Mendez,C. and Salas,J.A.
TITLE       Direct Submission
JOURNAL     Submitted (25-MAR-1998) Biologia Funcional, Universidad de Oviedo,
            Julian Claveria s/n, Oviedo, Asturias 33006, Spain
REFERENCE   4  (bases 1 to 15052)
AUTHORS     Quiros,L.M. Aguirrezabalaga,I., Olano,C., Mendez,C. and Salas,J.A.
TITLE       Direct Submission
JOURNAL     Submitted (13-SEP-1999) Departamento de Biologia Funcional e
            Instituto Universitario de Oncologia de Asturias (I.U.O.P.A.),
            Universidad de Oviedo, Oviedo 33006, Spain
REMARK      Sequence update by submitter
COMMENT     On Sep 16, 1999 this sequence version replaced gi:3023163.
FEATURES             Location/Qualifiers
     source          1..15052
                     /organism="Streptomyces antibioticus"
                     /strain="ATCC 11891"
                     /db_xref="ATCC:11891"
                     /db_xref="taxon:1890"
     gene            complement(590..1576)
                     /gene="oleW"
     CDS             complement(590..1576)
                     /gene="oleW"
                     /function="possibly involved in the C-2 deoxygenation step
                     in the dTDP-L-oleandrose biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="putative 3-ketoreductase"
                     /protein_id="AAD55450.1"
                     /db_xref="GI:5902166"
                     /translation="MPSPRLRFGVLGAADIALRRTVPALLAHPDVTVVAVSSRDTARA
                     ARFAAAFGCEAVPGHQALLDRDDIDALYVPLPVMVHTPWVEAALLRGRHVLVEKPLTA
                     TRSGAEDLIALARSRGLVLMENFTSLHHAQHGTVTDLLRDGTIGELRSLSAAFTIPPK
                     PEGDIRYQPDVGGALLDIGIYPLRAALHFLGPDLHAAGAVLRERRRNVVVSGHVLL
                     TTPHGVVAELAFGMEHAYRSEYTLFGTAGRLRLDRAFTPPETIHRPRVEIHRQDALDIV
                     DLPPDAQFANLVRDFVLAVREGPGRLTQHHADAVRQADLVERVMAVARVRWC"
     gene            complement(1596..3020)
                     /gene="oleV"
     CDS             complement(1596..3020)
                     /gene="oleV"
                     /function="involved in the C-2 deoxygenation step in the
                     dTDP-L-oleandrose biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="2,3-dehydratase"
                     /protein_id="AAD55451.1"
```

```
Db.  1147  CCCGGCTCGTTCTCGGCGACCAGCAGCGAGGGGCGAGATCTCCTCATGACCACCACATCCAT  1206
                  ||| ||  ||  ||||||  |||  ||  |||  | || ||  ||  |||| ||
Qy   3346  CCTCGTCACCGGCGGAGCCGGCTTCATTCGCTCCGCCTACGTCCGGCGGACCCTGCTGTCGCC  3405

Qy   1207  CCTGGTGACGGGCGGGGCCGGCTTCATCGGCGTCGCCACTACGTCGCGCAAACTCACCTACGGCCAGCCT  1266
                  ||  ||||||  |||||  |||  |||||||| ||  |||||||  || ||||||||
Db   1267  GCGCGGCGTGCCCGGACGTGACGGTCACCGGTCTCGACAAGCTGACCTACGCCGGGGACCCT  1326

Qy   3466  CGCCCGCCTGACGCGCCGGTGACGATCCCGGCCTCACCTTCGTCCAGGGGACGTGTG  3525
                  ||| || ||||  ||  | |||| ||  ||  ||  ||  |||||||| ||
Db   1327  CACCAACCTCGCCGAGGTGAGCGGACAGCAGCCGGTTCCGGTTCGCCGGGGCACATCTG  1386

Qy   3526  CGACACCGGCGGCTCGTCGACACGTGGCCGCCGGCGCAGCGACGCATCGTGCACTTCGGCGGC  3585
                  ||| ||  ||||| | ||||| || |  |||||  || |||  ||||  |||||| ||  ||
Db   1387  CGACGCACCCTGGTCGACGACCTGCTCCGCAGTGCACGACCAAGTGGTGCACTTCGCGCGC  1446

Qy   3586  CGAGTCGCACGTCGACCGCTCCATCACCCGACAGCGGTGCCTTCACCCCGACCAAGTGCT  3645
                  |||||||||||| ||||||  |||  |||  || ||  || |  || ||  |||| ||
Db   1447  CGAGTCGCACGTCGACCGTCCATTCTGGCCGCCGCCGACTTCGTGCGCCACCAATGTCAC  1506

Qy   3646  GGGCCACCCAGGTCCTGCGCCGCTCCGACACCGTGTGCCCACCTTCGTCGCACGT  3705
                  ||||||  ||||| ||| ||  ||| ||||| || | |||  |||  ||| |||
Db   1507  CGGTACCCAGCACGTCCTGCGACGCGGCGCGCCAGGGGATCGAGACGTTCGTGCACAT  1566

Qy   3706  CTCCACCGACGAGGTGTACGGCTCCCTCCCCGCACGGGCGCGCCGGAGAGCGACCCCCT  3765
                  ||||||||||||||||||||||  | |  | | || || | | |||||| |  ||||
Db   1567  CTCCACCGACGAGGTGTACGGCTGGCTCCATCGACGCCGGGTCCTGGGCCGCCGCCGT  1626

Qy   3766  GCTTCCGACCTCGCCGTACGGCGGCCTCGAAGGGCGCCTCATGGCGCTCGCCCA  3825
                  ||| || ||||  ||||  |||| |||  ||  |||  || |||||||| ||||
Db   1627  CAGCCCCAACTCGCTCTATTCCGACGCCCAAGGGCGTCCAGCGATCTCCGTGGCCCTA  1686

Qy   3826  CCACCGCACCCACCGGCTGGACGTCGGACGTCGGGGTGAACCCGTTGTTCGAACAACTTCGGCCCCCCA  3885
                  |||||||||||||||  ||| ||||| ||||||||  |  |  ||  ||  ||| ||  || |||| |
Db   1687  CCACCGCACCCACCGGTCTGGACGTCTGGACGTCGGGGTTGCTCAACAACTACGGGTCCCA  1746

Qy   3886  CCAGCATCCCGAGAAGCTCATACCCGCCTTCCTGACCAGCCTCCTCGTCCCGGCGCACCGT  3945
                  |||   || || ||||| |||  ||| | |  |||||||| |  |||  |||||| ||||
Db   1747  TCAGTTCCCCGGAGAAGGTCATCCCCGTGTTCGTCACCAGCGTCGTGCTGCCTCCGGAGT  1806

Qy   3946  TCCCCTCTACGGCGACGGGCGCACGTGGCTGCGGCGACGTGGCTGCACGTCGACGACACGTCAG  4005
                  ||  ||  ||||||| |  ||||||||  ||| | |||| | ||||||||||  ||||||||
Db   1807  GCCGTTGTACGGGGACGACGGACACCAACGGCTACCGGCGCCACGTCTTGCTGGAGACCCCG  1866

Qy   4006  GGGCGTCGAACTCGTCCGCGTGTCGCCCGTGCTGGGCGGGGAGATCTACAACATCGGGGCGGCGG  4065
                  |||| ||| |  ||| ||||  | ||  ||| ||  || ||| || |||  ||||||| || |||
Db   1867  CGGATCGACGTCGGTCCGCACGGGACGTCGGGCGAGGTCTACAACATCGGGCGCGGCGCG  1926

Qy   4066  CACCTCGCTGCCCAACCTGGACGTTGCTCGCAGGTTGCTCGCACGTGCGCGGGGCGGGGCC  4125
                  ||||| || || |||||| |||| | ||  ||| | || ||  || |  || ||| || ||
Db   1927  CACCGAGCTGTCCAACAAGGACGCTGACGCCAGCTGACCTCCTGGAGCAGCCGTGTGGCCCGCTG  1986

Qy   4126  GGAGGCCGCATCGTCCACGTCGAGAACCGCGAGAACCGCCAAGGGCACGAGGGCACCGGCTACGCGGTCGACCA  4185
                  ||||| ||| || |  ||||||||||| | |||| | | |||||  | |  |||| | ||||||||||
Db   1987  GGACCGGTGTACGGCGGGCTAGGTCACCGTCACCGGCCGGATCACCGTCAGCGTGCTGGTCTCCGGCGACTG  2046

Qy   4186  CAGCAAGAATCACCCGGGAACTCGGTTACCCGGACCTCGGCGACCGGCGCTGGC  4245
                  ||||| ||  ||||  ||  |  |||  |||||| |||||  |||||| || ||
Db   2047  CACGAAGATCCGGCCGTGAGCTCGGGGTACCCGGACCTCGGGAGTCGGGACGCACTGGC  2106

Qy   4246  CGACACCCGGAAGTGGTACGACGAGGACTGGTGCGGCTGGTGTCCCCTG  4294
                  |||| | ||  | | || || |||  | |  ||||||
Db   2107  CGAGACTGTCGCCGTCTGGTACCCGCACCACCGTGCTGGGAGGCCCGCTG  2155
```

```
RESULT 6
AF055579
LOCUS       AF055579   15052 bp   DNA   linear   BCT 06-JUL-2000
DEFINITION  Streptomyces antibioticus putative 3-ketoreductase (oleW),
            2,3-dehydratase (oleV), dTDP-4-keto-6-deoxyglucose 3,5-epimerase
            (oleL), dTDP-D-glucose synthase (oleS), dehydratase (oleE),
```

```
                     /db_xref="GI:5902167"
                     /translation="MIWGIPAMSEAMGSVPTAGSEVSSTCAFLSWLDARRRANRLTVE
                     HVPFRELSQWQFDENTGNLRHTSGRFFSIEGLRVRTDHCWTGHSTWQPIIVQPEIGILG
                     LLVKRFDGILHVLVQAKNEPGNIGGLQLSPTVQATRSNYTRVHRGGGVRYLEYFASPR
                     GRGRVLADVLQSEQGSWFLHKRNRNMVVEALDDVPLDDDFHWISLGGLRKLLLRPHLV
                     NMDTRTVLSCLPPDPAPDGRQPPAPAAPFAAAVTRSLTRGATALHTMGEILGWLTDER
                     SRRELVQQRVPLEETAFSGWRRDHAIAHKDGYFRVIGVSRASSREVSSWSQPLLA
                     PVGPGLAAFVTRRIRGVLHVLLHARTEAGLLNGPEMAPTVQCRPLNYRAVPAEYRPAY
                     LDYVLSADPGRIRYRDTLQSEEGGRFPHHAENRYVVVEAEDDFPVEVPRDFRWLTLHQIL
                     ALLHHSNTVNVEARSLVACIQALS"
gene            3356..3970
                     /gene="oleL"
CDS             3356..3970
                     /gene="oleL"
                     /function="epimerization at carbons 3 and 5 in the
                     biosynthesis of L-oleandrose"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-4-keto-6-deoxyglucose 3,5-epimerase"
                     /protein_id="AAD55452.1"
                     /db_xref="GI:5902168"
                     /translation="MELLDVDGAWLYTPEIMRDERGEFLEWFRGRTFQEKIGHPLSLA
                     QANCSVSRRAFCAASTSPPPPGQAKYVTCASGVELVVDPLDVFVDFGPTFGRWAAVRLD
                     AARHQGLYLAEGLGHAFMALTDDATVVYLCSQPYVAEAERAVDPLDPAIGIEWPTDID
                     IVPVGEGTPTHRPWRRPRRPGILPDYEGVPGALHRGGHRGTGP"
gene            3967..5037
                     /gene="oleS"
CDS             3967..5037
                     /gene="oleS"
                     /function="activation of glucose-1-phosphate into
                     dTDP-D-glucose"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-D-glucose synthase"
                     /protein_id="AAD55453.1"
                     /db_xref="GI:5902169"
                     /translation="MKALVLAGGSGTRLRPITHTSAKQLVAVANKPVLFYGLEAIAAA
                     GITDVGLIVGDTAGEVPRAVGDGAKFGLDITYIEQSRPLGLAHAVLIAHTYLGDDDFV
                     MYLGDNFIVGGIDDLVRTFRDGRRPAARILLTHVSPSGFGVAELDDDGRVGLEEKP
                     RHPKSDLALVGVYFFTPAIHEAVRAIEPSWRGELEITHAIQHLIDNGADIQSMVIEGY
                     WKDTGNVADMLEVNRTVLEDLEPRIEGTVHEADYFRVIGVSRVTNSRIMGPAII
                     GAGPEISDSYIGPFTSVGDNCRITGSEMEFSIMLAESAITGVRRIEGSLIGRNVQVTQ
                     SLHAPNAHRFVLGDHSKVEIQS"
gene            5034..6041
                     /gene="oleE"
CDS             5034..6041
                     /gene="oleE"
                     /function="4,6-dehydration during 6-deoxysugars
                     biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="dehydratase"
                     /protein_id="AAD55454.1"
                     /db_xref="GI:5902170"
                     /translation="MNLLVTGAAGFIGSRYVHHLLEATRRGREPAPVITVLDKLTYAG
                     VLGNVDDPAVTFVRGDIADAPLVDSLMAEADQVVHFAAETHVDRSITSPGTFVRTNV
                     LGTQVLLDDAALRHGVGPFVHVSTDEVYGSIEHGSWPEHQPLCPNSPYSASKASSDLLA
                     LSYRTHGLDVRVTRCSNNYGPHQFPEKTVPLFVTNLLDGLRVPLYGGDGLNVREWLHV
                     DDHCLGVDLVRTQGRPGEVYHIGGGTELTNRDLTGLLLDAFGVGWDVVDPVADRKGHD
                     RRYALDCAKAADELGYRPRRDFAEGIARTIDWYRDNRAWWEPLKKRPAGPAAPPRGSG
                     P"
gene            6038..6925
                     /gene="oleU"
CDS             6038..6925
                     /gene="oleU"
                     /function="involved in L-oleandrose biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="4-ketoreductase"
                     /protein_id="AAD55455.1"
                     /db_xref="GI:5902171"
                     /translation="MRWLITGAAGMLGRELVRRLAENEEDVAALGHDHLDVTRPSAVR
                     AALAEHRPGIVVNCAAYTAVDDAETDEAAAALLNAEAPRILLAEGLRPHRRHGLVHLST
```

```
                     DYVFPGDARTPYAEDHPTAPRSAYGRTKRDGEQAVLTALPTATVLRTAWLYGRTGRSF
                     VRTMIEREARGGAIDVVADORGQPTWTGDLADRIIAVGRHPGVHGILHATNAGSATWY
                     DLAQEVFRLLDADPGRVRPTTGAAFRRPAPRAYSVLGHDRWRGTGLAPLRDWRSALR
                     EAFPDILAAEHPPTRRGAA"
gene            6922..8103
                     /gene="oleNI"
CDS             6922..8103
                     /gene="oleNI"
                     /function="possibly involved in the transamination,
                     together with oleN2, during the biosynthesis of
                     D-desosamine"
                     /codon_start=1
                     /transl_table=11
                     /product="dehydratase"
                     /protein_id="AAD55456.1"
                     /db_xref="GI:5902172"
                     /translation="MKRGVHDLALFGGDAAFLQPLYMGRPNTGDRKRLLDRLEWALDN
                     RWLTNGGPLVREFQRIADLAGVRNCVATCNATAGLQLLLREAEVTGEVIMPSMTFVA
                     TAHAVRWLGLRPVFCDIDPDTGCLDPKLVEAAVTPRTGAILGVHLWGRPSRVDELAAI
                     AAEHGLKLFYDAAHALGCTSRQRRLGSFGDAEVFSFHATKVVNSFEGGGIVTDDDTRA
                     ERLRALHNFGLGHDGVGAINAKMSEAAAMGLTSLEAFADAVASNRANYELYRQELS
                     GLPGVRLIDYDPAERNNYHYVIALIDAGVTGLHRDLLLTLLRAENVVAQPYFSPGCHQ
                     REPYRTEHPVSLPHTEHLAEQVIALPTGPAVSREDIRRVCDIIRVAAAHGPRITAQAG
                     A"
gene            8225..9682
                     /gene="oleT"
CDS             8225..9682
                     /gene="oleT"
                     /function="possibly responsible for the 3,4-reduction step
                     in D-desosamine biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="reductase"
                     /protein_id="AAD55457.1"
                     /db_xref="GI:5902173"
                     /translation="MALQVDRTPLDPIAVCAWPGGNSAAAAALRPLIEEDAPTGLEP
                     DRLAEHLIALARRYGTEPFTPLESARRGLGLDRATFARVLAVFHRTPALRTAVERGPA
                     GMYWTNTILPLERRGVLDAAVRGEPAFPYSVGLYPGSCMFRCHFCVRTGARYQQSA
                     LTDGNAMFASLIDRMPTDNPHALYLSGGLEPLTNPGTGDLVRRAAARGFKLSLYTNSF
                     ALTRQTLDRQPGLWDLVALRTSLYGLSEDDYVATTKKGAFQRVKDNLTRFQALRRER
                     RRPVRLGLNYIILPGRAGRLTGLADYIGDLNDGGPDRPVDFLTLREDYSGRPDGKLAP
                     VAVQVDLLGDVYLYREAGFPGLQGAERYVAGRLTTGTELEEVVRRFVTEGRQVAPRPG
                     EEYFLDGFDQTVTARLNQMETDIADGWAEHRGFLR"
gene            9957..11231
                     /gene="oleI"
CDS             9957..11231
                     /gene="oleI"
                     /function="intracellular inactivation of oleandomycin by
                     glycosylation"
```

```
Query Match          14.2%;   Score 973.4;  DB 1;  Length 15052;
Best Local Similarity 67.7%;   Pred No. 9.4e-66;
Matches 1396;   Conservative   0;  Mismatches 656;   Indels   9;   Gaps   2;

Qy  2268 GTGAAGGCCCTGGTCCTGCCAGGTGGAACCGGCCAGCAGACTGAGGCCGTTCACCCACACC 2327
         ||||||||||||||| || |||||| ||| ||| ||| ||||| || ||| |||||||||
Db  3967 GTGAAGGCCCTGCTACTGGCCGGGTTCCGGCGACACCCGGCTCGCGGCCGATCACCCATACC 4026

Qy  2328 GCCGCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTACGGCGTCGGAGTCC 2387
         || ||||| || |||  ||  ||| ||| |||||||||| |||||  | ||||| |||||
Db  4027 TCCGCCAAGCAACTGGTCGCCGGAGGCCGGCGACAAAACCGGTGCTCCTTCTACGGCGTCGGGAGGCC 4086

Qy  2388 CTCGCCGCGCGGCGGGTGCCGGGAGGCCGGCGTCGTGGGCCGCGTACGGCCGCGGGAGATC 2447
         ||  ||| ||| ||||| | | |||||| | ||||  || ||| ||||||| | | |||
Db  4087 ATCCGGCCGCCCGCGGGGGATCACGAGCTCGGCCTGATCGTCGGCGACACCGGCCGGGCGAGGTC 4146

Qy  2448 CGGCAACTCACCCGGCGCACGGGCCGTTCGGGTTACGGCCATCACCTACCTCCACCAGCCC 2507
         | |||| |||||||| ||| ||||  ||| ||| ||||  |||| ||||| || ||||||
Db  4147 CCGCGGCCGCCGCCGGCGCACGGCGCCAAGTTCGGCCTCGACATCACCTACCTCACCAGCAGC 4206

Qy  2508 CGCCCGCTCGGTCGCCGACGGTGCGCCATCGCCCCGCGGCCTTCCTCGGCCGACGCAGCAC 2567
         ||| || ||||| || ||||| || ||  ||| |  || ||| || || ||||||||||
```

The following is a nucleotide sequence alignment (query "Qy" versus database "Db"), with position coordinates on each line.

```
Db  4207  CGCCCCTGGGCTCGCCCACGCGTACTGATCGCCCACACCTACCTCGGCGACGACGAC  4266
Qy  2568  TTCCTGCTGCTACCTGGGGGACACTACCTGCCCCCAGGCGTCACCGACCTTCGCCGCCAA  2627
Db  4267  TTCGTCATGTAACCTGGGCGACAACTTCATCGTTCGGCGCCTGCTCGCCCGGCGACCGTCC  4326
Qy  2628  T---CGGCGCCGATCCCGGCGGCGGCCCCGCCCGGATCCTGTCACCCATGTGTCCGACCCCAGC  2684
Db  4327  TTCCGGACGGCAGGCGGCCGCCCGGCGATCCTGTTCACCCCATGTGTCCGACCCCAGC  4386
Qy  2685  GCCTTCGGCGTCGCGGAGTCGACGCGGAACGTGCTCGCGCTTGGAGGAGAAACCC  2744
Db  4387  GGCTTCGGCGTCGGCGAACTCGACGACGACGCCGTCGCCGTCGCCTCGGAGGAGAAGCCC  4446
Qy  2745  GACGTCCCGCCAGCTCGCTCGCCGTCATCGGCGTGTACGCCCGTTCAGCCCGGCCTCCAC  2804
Db  4447  CGGCACCCGCAAGAGCGACCTGGCACTGGTCGGCGTCTCTTCACCCCCCGCCATCCAC  4506
Qy  2805  GAGGCGGTACGGGCATCACCCCCTCCGCGCGGGAGCTGAGATCACCCACGCCGTG  2864
Db  4507  GAGGCCGGCCCGTTCACCATCGAAACCCTCCGGCGGGCGCGCGGGAGATCACCACGCCCATC  4566
Qy  2865  CAGTGGATGATGATCGACACCGGGGGCTCGCGCTACGGGCGACCACCACCCGGCGCGC  2924
Db  4567  CAGCACCTCATCGACAACGCGCCGACATCCAGTCCAGTCATCGGTCATCGAGGGGCTACTGGAAG  4626
Qy  2925  GACACCGGCAGCGCGGAGGACAATGCTGGAGGTCAACCGTCCTGGAGGTCGCACTGGAG  2984
Db  4627  GACACGCCAAACGTCGCCGACACGTGCTGGAGGTCAACCGGACCGTACTGGAGGGACCTGGAG  4686
Qy  2985  GGGCGCATCGAGGGGAAGGTCGACGCGCCACAGCCACACCTGGTCGGCCGGGGTCCGGGTGCC  3044
Db  4687  CCCCGGCATCGAAGCCACCGGTGAGCAGCACCCGTCGGTCGTGGTCGCAGGTGTCGGTGGTCGGA  4746
Qy  3045  GAAGGCGGCATCGTGCGGGTCTCACACGTTGCGGGGGTGGTGATCGGTGGTGGTGGTGCC  3104
Db  4747  GAGGGCGCGCGGGTGACGAACTCGCGCATCATGGCCCCGCCCATCATCGGCGGCCGCGCCCCCCC  4806
Qy  3105  GTCGTCAGCAAATCCAGTTCGTCGCCGCCCGTACACCTTCCATCGGGAGGACTGCCCGGGTCGAG  3164
Db  4807  GAGATCTTCCGACTCCTACATCGGCGTCGTCACCTGCGGCGTCGGCCGACAACTGCCCATCACC  4866
Qy  3165  GACAGCGGCCATCGAGTACTCCGTCGTCCTGCGGGGCGCCAGTGAGGGGCGTCCCGC  3224
Db  4867  GGCGACGAGATGGAGTTCTCCATCATGTCGTCCCATCCAGTCGGCGATCACCCGCGTCGCCGCT  4926
Qy  3225  ATCGAGGGCGTCCCTCATCGGCGTGCTACACCTCCCCGGCCCCGTCTCCCCCGCAG  3284
Db  4927  ATCGAGGGGGGATCACTGATCGGCAGGAACGTTCCAGGTGACCAGCGAAACGCTGCACGCCCCCAAC  4986
Qy  3285  GCTCACCGACTGGTGATCGGCGACCACAGCAAGGTGTATCTCACCCCATGACCACGACCA  3344
Db  4987  GCCCACCGCTTCGTCGTCGGCGAGCGGGTTCATCGGCGACCAAGGTGGAGATCCAGTCAT-----GAACC  5040
Qy  3345  TCCTCGTCACCGGCGAGCGGGTTCATCGGCTCCGCTCCACGTCGCCGCCTCCGGCTCCTGTCGCG  3404
Db  5041  TCCTGTCACCGGCCCCGGATCATCGGCTCACGTCACGTCACCTGCTGGAAG  5100
Qy  3405  CCGGGCCCCGGCCCCGCGGTCGGCGTCGGCGACCTCTCGACAAACTCACCTACGCCCGGCAGCC  3464
Db  5101  CGACGCGGCGGCGGCGGAGCCCGGGGAGTCGGTGATCACGGTGCTGGACAAGCTCACCTACG  5160
Qy  3465  TCGGCGCCTGCACGCGGCTGCGTCGACCATCCGGCCTCACTTCGTCCAGGGCGACGTGT  3524
Db  5161  CCGGGCGTCCTCCGGCAACGTCCCGACGACGCCCGCCGTCACCTTCACCTTCGTGCGGGGTCACATCG  5220
Qy  3525  GCGACCACCGCGCTCGTCGACACGCTGGCGCGGCATGCGTCGACATCGTGCACTTCGCCG  3584
Db  5221  CCGAGGCCACCAGTGGTGGGACTCCCGTGGGACTCCCGACTGGTTCGCACTTCGGCCG  5280
Qy  3585  CCGAGTCGGCCACGCTGCAGACCCCGCTCACCATCACCCGACCAGCAGCCGCCCACCGTC  3644
Db  5281  CCGAGACCCACGTGGACACCCGGCGACCCTCCAATCACCTCCGCCGGGCGAACCTTCGTCCGCCACCAACGTCC  5340
```

```
Qy  3645  TGGCACCCAGGTCTCGCTGCTCGACGCCGCCACGCCACCTGTGCCCACCTTCGTGCACG  3704
Db  5341  TCGGCACCCAGGTGCTCGCGACGCCCTGCGCCGCACGGCGTCGGCGCCGTTCGTGCGCACG  5400
Qy  3705  TCTCCACGCAGGAGTGTACGGCGTCACGGCGTCTACGGCGTCCCGGGACCCGTCC  3764
Db  5401  TGTCCACGGACGAGTCTACGGCTCCCATCGAGCACGAGATCGGCGGATCCTGGCCGGCCAGCCGC  5460
Qy  3765  TGCTTCCGACCTCGCCGTACGGCCGCGTCGAAGGCGGCGTCGGACCTCATGGCGTCCGCCC  3824
Db  5461  TGTGCCCGAACTCGGCGTACTCCGCCTCCAAAGCCTCGACCTGCTCGGCGTCCT  5520
Qy  3825  ACCACCGCACCCACACCGCCTGGAGCTCCGGAGCTGCGTGTTCGAACAACTTCGGCCCCC  3884
Db  5521  ACACCGCACCCACGCCCTGGAGCGTCCGCGCTGCTCCAACAACTACGGGGCCGC  5580
Qy  3885  ACCAGCATCCCGAGAAGCTCATACCGCGCTTCCTGACCACCCTCCTGTCGGCCGGCACCG  3944
Db  5581  ACCAGTTCCCCGAGAGATCGTCCCGCTGTTCGTCACCAACCTCCTCGACGGCTCTGCGGCC  5640
Qy  3945  TTCCCCTTACGGCGACGGGCGCGGGAGGACCCACCACGCCCTGGCC  4004
Db  5641  TCCCCCTCTACGGCGACGGCCTCAACGTCCGCGAGTGGGTCGCACGTGGACGACCACTGCC  5700
Qy  4005  GGGGCCGTCGAACTCGTCCGCGTGTCGGCGGCGGCCGGCGGCGAGAGATCACATCGGGGGCG  4064
Db  5701  TGGGCGTCGACCTGCTCCGACCTGGTCGCCACCGTTGCTCGCACTGTGCGGCGCGGGGCC  5760
Qy  4065  GCACCTCGCTGCCCAACCTGAGCTCAGGCCGACCTTGCGCTGTGGACGCGTTCGGGGCT  4124
Db  5761  GCACCGAACTGACCAACCGGGACCTTCGGGGCGCCGCGCGCTACGCCCTACGCCCCGACT  5820
Qy  4125  CGGAGGCCGCATCGTCGAGAACGTCGAGAAGGGGCCAAGGGCACGGCACGACCACGACTCGGGGCGACCGGCTGG  4184
Db  5821  GGGACGTCGTGGACCCGGTGCGGCGTACGGCACGCAGCGCACGCAGCGCACGACCACGACGCATCG  5880
Qy  4185  ACAGCAAGATCACCCGGGCGGAACTCGGTCGTACCGGCCGACCGGCTGG  4244
Db  5881  GCGCCAAGGCCGCCGACGAGCTCGGCGTACCGGCGCACCGAGGGCATCG  5940
Qy  4245  CCGACACCGGCGAAGTGGTACGAGCGGCAGCCGCGGACTGGTGCGTCCCCTCCGCCGA  4304
Db  5941  CCCGACCACCATCGACTGGTACCCGGTACGGTTGGGAACCCCTGGTGGTGGGAACCCCTGTGAAGAAGCGGGC  6000
Qy  4305  CATGACGTCGGGCCGGACCCG  4325
Db  6001  CGGCGCGGTCCGCGCCGCCCCCC  6021
```

RESULT 7
SGAJ6985
LOCUS       SGAJ6985            25459 bp    DNA     linear   BCT 10-MAR-2001
DEFINITION  Streptomyces glaucescens GLA.0 gene cluster for
            5'-hydroxystreptomycin biosynthesis.
ACCESSION   AJ006985
VERSION     AJ006985.1  GI:3256042
KEYWORDS    sph gene; strA gene; strB1 gene; strB2 gene; strD gene; strE gene;
            strF gene; strG gene; strH gene; strI gene; strK gene; strL gene;
            strN gene; strO gene; strP gene; strQ gene; strR gene; strS gene;
            strT gene; strU gene; strV gene; strW gene; strX gene.
SOURCE      Streptomyces glaucescens
ORGANISM    Streptomyces glaucescens
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 25459)
AUTHORS     Retzlaff,L., Mayer,G., Beyer,S., Ahlert,J., Verseck,S., Distler,J.
            and Piepersberg,W.
TITLE       Streptomycin Production in Streptomycetes: a Progress Report
JOURNAL     (in) Baltz,R.H., Hegeman,G.D. and Skatrud,P.L. (Eds.);
            INDUSTRIAL MICROORGANISMS: BASIC AND APPLIED MOLECULAR GENETICS:
            183-194;
            ASM Press, Herndon, VA, USA (1993)

```
REFERENCE   2  (bases 1 to 25459)
  AUTHORS   Beyer,S., Distler,J. and Piepersberg,W.
  TITLE     The str gene cluster for the biosynthesis of 5'-hydroxystreptomycin
            in Streptomyces glaucescens GLA.0 (ETH 22794): new operons and
            evidence for pathway-specific regulation by StrR
  JOURNAL   Mol. Gen. Genet. 250 (6), 775-784 (1996)
  MEDLINE   96204519
   PUBMED   8628239
REFERENCE   3  (bases 1 to 25459)
  AUTHORS   Piepersberg,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Piepersberg W., Institut fuer Chemische
            Mikrobiologie/ FB 9, Bergische Universitaet Wuppertal,
            Gauss-Strasse 20, 42097 Wuppertal, 42097, GERMANY
FEATURES             Location/Qualifiers
     source          1..25459
                     /organism="Streptomyces glaucescens"
                     /strain="ATCC 13032 (GLA.0)"
                     /strain="ETH 22794 (GLA.0)"
                     /db_xref="taxon:1907"
     gene            complement(14..1789)
                     /gene="strW"
     CDS             complement(14..1789)
                     /gene="strW"
                     /function="putative transport of
                     5'-hydroxydihydrostreptomycin phosphate"
                     /codon_start=1
                     /transl_table=11
                     /product="StrW"
                     /protein_id="CAA07371.1"
                     /db_xref="GI:3256043"
                     /db_xref="SPTREMBL:Q54204"
                     /translation="MSRTRLPVAGTAEVRAHGAALARRHARDLTWMiALHGLAAVCGL
                     LGPWLIGRLLDEVVRGTTLSTVDHIGTALLLSLVAQAVLTYWAVQRSCRFGEKVTAEV
                     REDFVDRVRLPLPPTVEDAEPGDLITRASRDTDALTNTVRYGVPETLIALMTCLFTFA
                     ALVLVGPLTALPSLVVVPLLWAGTRWTLRRSQAAYRRGASYTALGDTLAEFTVPGART
                     VEALRLQAWRRALDRDLAEAYEAERHTMCWRSSWYLTVELSYVVPVVATLAIGGLIY
                     TOGLARVGQAVAATLVVRQLIFPLDDVLSWMDELQIGSASLARVLGVSELPDDRRPSG
                     ARPDGROLTARSAGYSYVPGRDILHGVDLDLAPGERLAVVGPSGSGKSTLGKILAGVY
                     EPVSGVVELGGVRLVDLGTEELRRRIALVTQDHHIFRGHLRENVTLARPAADDEVEHA
                     LKAVGAAEWAAALPDGPETVVGSGHOLTAAQAQQIALARLVLADPHTVVLDEATALV
                     DPATARHTERALAAVLSDRTVIAIAHRIHTAQSADRIAVLENGRIKELGSHOELLAAN
                     APTPPCGAYGTAIELIAEIPRPVLA"
     gene            complement(join(1786..3540,3547..3552,3572..3577))
                     /gene="strV"
     CDS             complement(1786..3540)
                     /gene="strV"
                     /function="putative transport of
                     5'-hydroxydihydrostreptomycin phosphate"
                     /codon_start=1
                     /transl_table=11
                     /product="StrV"
                     /protein_id="CAA07372.1"
                     /db_xref="GI:3256044"
                     /db_xref="SPTREMBL:Q54203"
                     /translation="MRPLPAADPGNPDIRSAARFLLWLAGRIAWPLSAAAFCGVLWLG
                     SQALMPVVIREAIDDGLVPGDMDAVLSWSAVLLGLGVVQAVSGTVRRRLNMYNSLSAG
                     FRTVQLVGHANTLGATLERRVDEGELVSVGTSDMNALGHAFDIVGRTVGAVISVVLV
                     SVVLVRQSATLGMMLIVAVPAIVLTMGYTLRPLHRRQSAYRDLQGELTSRLTDVVAGL
                     RVLRGIGGETPFARSYRERSQRVRAAGARVAGAEALIRAAQVLLPGLLVGTVTWIGAR
                     AALDGSVPGELVAFYAYAAFLVEPLGIFTETADRPARAHYAARRVVTVLRMRPKARP
                     EHADGPMGTLATALVDETTGLIVTPGTHVVAPSAEGVRLADRLGGYDEAPVRHDGR
                     LLTALAPDEVRRRILVVDNRPRLFRGPLRASLDPHVTADDTAITAALRARAAEDIVGL
                     LPDGLDTDITADGRHFSGGQRHLVLARALLADPDVLIVVDPTSAVDAHTEDLIGRRLR
                     AARSGRTTVLVSSSPLLNQADRVSFLSGGRVAAAGTHRELLAASEDYRLVIQRETGE
                     APPGPLPSTDLVTRGSAE"
     -10_signal      complement(3547..3552)
                     /gene="strV"
                     /citation=[2]
                     /evidence=experimental
     -35_signal      complement(3572..3577)
                     /gene="strV"
                     /citation=[2]

     gene            3587..3615
                     /gene="strV and strX"
     protein_bind    3587..3615
                     /gene="strV and strX"
                     /note="The expression of both the genes, strX and strV, is
                     activated by binding of the protein StrR"
                     /citation=[2]
                     /bound_moiety="activator"
                     /function="StrR binding site"
                     /evidence=experimental
     gene            3706..4254
                     /gene="strX"
     CDS             3706..4254
                     /gene="strX"
                     /function="putative 5'-hydroxystreptomycin biosynthesis
                     protein"
                     /codon_start=1
                     /transl_table=11
                     /product="StrX"
                     /protein_id="CAA07373.1"
                     /db_xref="GI:3256045"
                     /db_xref="SPTREMBL:Q54202"
                     /translation="MLIHETPLAGAALIELSELRDERGFSRSFCREEFERAGLEPAV
                     VQCNLSFNHRAGSLRGFHYQTEPNAEAKTIRCTRGAIYDVIVDLRPDSPTFLRHFGAE
                     LTADNRLALHVPRNFAHAYLTLADDTETVYQVSAAYTPGAERGLRWNDPALGVEWPAP
                     VKLVSDKDAGWPLLSEVPAGSR"
     gene            4251..5537
                     /gene="strU"
     CDS             4251..5537
                     /gene="strU"
                     /function="putative 5'-hydroxystreptomycin biosynthesis
                     protein"
                     /codon_start=1
                     /transl_table=11
                     /product="StrU"
                     /protein_id="CAA07374.1"
                     /db_xref="GI:3256046"
                     /db_xref="SPTREMBL:Q54201"
                     /translation="MIIVDTALRREERDRPIRVGMVGAGFMGRGLARHIVRSVPGMR
                     LAAIANRSLATAERAYTEAGCVRPVRANGAGDVEAAVAAGRPVTEDAFALLAAEGIDC
                     LVDVTGAVEFGARVTVAALERGLPVVTMNAELDGTVGPLLAHRARAAGVLTAADGDQ
                     PGVGQNLLRFVRGLGVTPLVAGNVKLQDEYRTPATQKAFAERWGQNVHMVTSFADGT
                     KVSFEQAIVANAHGLTVARRGMYGRHTGHVDELTGVVYDIDELRELGGVVDYVVGARP
                     GPGVVIGTHDDPRQRHYLELYKLGKGPLYSFYTPYHLCHFEVPHTIARAVDFADAAL
                     TPPAGPRVDVVATAKRDLKAGTVVDGIGGYDTYGVAETHAATLASRLLPMGVAEGCVL
                     RRDIARDAVLTYDVRLPAGRLVDRLREEQNELFPG"
     gene            complement(5455..6354)
                     /gene="strQ"
     CDS             complement(5455..6354)
                     /gene="strQ"
                     /function="alpha-D-glucose-1-phosphate
                     cytidylyl-transferase"
                     /citation=[2]
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="StrQ"
                     /protein_id="CAA07375.1"
                     /db_xref="GI:3256047"
                     /db_xref="SPTREMBL:Q54266"
                     /translation="MTVDSDIPVVLLCGGYGTRIREASERLPKAMVDIGGRPILWHIM
                     KIYSHYGFRRFVLCLGYKGWAIKQFFLDYRAQLSDFTISLSDEHRIRFHDDRADENCE
                     ITFAETGLDTATGAARLRVRRYVDTTDDFMMGYGDGVGIDLAALEKEHRAAGRIGVT
                     GIHPTSKFGEMDVEDDQVLEFNEKPTRVTGFVSGGFMFRREFLDDYLDDDPKLWLES
                     APLQRLARDRQLTLHPHHDFWCAMDTFKDYQHLNELWAAGDAPWKVWEQPGNRSVSPG
                     TARSAPRAAGRPAGPPAAVVRHR"
     gene            complement(6351..7427)
                     /gene="strP"
     CDS             complement(6351..7427)
                     /gene="strP"
                     /function="putative protein involved in biosynthesis of
                     5'-hydroxystreptomycin"
                     /codon_start=1
```

```
            /transl_table=11
            /product="StrP"
            /protein_id="CAA07376.1"
            /db_xref="GI:3256048"
            /db_xref="SPTREMBL:Q54262"
            /translation="MRILLTGHQGYLGTVMAPVLTAAGHQVTGLDSGLFASRVLGALD
            VPDVPGLALDLRDVTVDTLRGYDAVVHLAALSNDPLGSLDPELTHAVNHHASTRLARL
            AKEAGVTRPAYASTCSVYGAQSGDAPVDEDAPLKPVTPYAVSKVRVEDDLVELADSGF
            APFFLRNATAFGFSPRLRSDIVLNNLVGRAVLGGRVTVLSDGTAWRPLVHAEDIARAV
            LAGLAAPAATVRARAFNIGTEENRTVAEIAEAAAAVPGSVVEISEGTGNDPRSYRV
            DFTRARTELGHRATWTIPDGAOLAREYRARGLTRAAFAEHDFTRLAVLERRRREGSLD
            GDLRVRPARVRRDGDHRTATAAEG"
gene        complement(7522..8679)
            /gene="StrH"
CDS         complement(7522..8679)
            /gene="StrH"
            /function="putatively involved in biosynthesis of
            5'-hydroxystreptomycin"
            /codon_start=1
            /transl_table=11
            /product="StrH"
            /protein_id="CAA07377.1"
            /db_xref="GI:3256049"
            /db_xref="SPTREMBL:Q54261"
            /translation="MQTSAHPAPARAADPSYEEALLSQDERFRTFLAGLPADTPVWFC
            FFSSGLMHVIQHFLRFVPPRLNIAFVCSGLTDDEIRLRDTMSGGRPAPFLDERVGSHE
            IFEFLLRNLDRPEFGIVDSDCFWADWFWRCTDDLREPDVAVSSPLSVGPIPLAAPPEM
            AINPLARPEIERAIGGPVSPAAISYTPPGPEKEIDGAMVRLIEPHHEAALAKVLAMEG
            PRLPFPQGGLLDVLDDGREVRSHERHHHAQNGNGVIRVVFDGLMFYQLMALAAGWRIA
            HFHSFPGTKVFSAOMVHAGGISTWHRLRPSEIRAGIHELPWSAHIDALLLENFTARPD
            APPAYRARGTRLAANLRQSCTTLGRLRRDDLRGKLASSGVDVTDPRWRAVLG"
gene        complement(8802..9407)
            /gene="strG"
```

```
Query Match    13.28;   Score 903.2;   DB 1;   Length 25459;
Best Local Similarity   64.58%;
Matches 1444;   Conservative   0;   Mismatches   738;   Indels   58;   Gaps   4;
```

```
Qy   2244 CGACCCGCTCCGCGGAGGCCGAGGTGAAGGCCCTGTCCTGGCAGGTGGAACCGCAGC 2303
               | ||      |  | ||||||     | | ||||||   ||
Db  18772 CGTCACGTCACGCAGGGAGAGAACATGAAGGCGCTCGTCCTGGCCGGCGACCGACCC 18831

Qy   2304 AGACTGAGGCCGTTCACCCACCACCACCAGTCCCCGGAAACAGCAGCTGCTCCCCATCGCCGCCAACAAGCCC 2363
               | |||||  ||||||| ||||||||||     |     |||   | ||||||
Db  18832 CGCTTACGGCCGATAACCCACCACCACCGTCCCGGGAAACAACTCGTCGCCGCCAACAAGCCG 18891

Qy   2364 GTGTCTTCTACGGCGTGGAGTCCCTCGCCGGCGGTGTCCGGGAGGCCGGCGGCCGTCGTC 2423
               | ||||| | ||||||||||||||| | |  |||      |||     |||| ||
Db  18892 GTCCTCTTCTACGGCGGTGGAGTCCCTCGCCGGACTCGCGGGCCGGCGGGATAACCGATGTCGGCCATCGTG 18951

Qy   2424 GTGGGCGGCGGTACGGCCCGGAGATCCGCGGAACTCACCGGCGCACCGCGTTCGGGTTA 2483
               | ||  |  |||||||||||||       |     |||      |||     ||||
Db  18952 GTCGGGGACGCCGTCGCGGAGATAACCGAAGCCGTCGGGGGACACGGCTCGAAATTCGGTCTC 19011

Qy   2484 CCGATCACCTACCTCCACCAGCCCGCGCGAAACCACTCGGCCTCGGGGGACGCTCCGCATTTCC 2543
              |||||| | |||||||||||||||| |||||||||||||||||     || |||||||
Db  19012 GACGTCACCTACACATCCCCAGCGAAACCACTCCCAGCGAAACCACTCGGCCTCGGCACGCGCACGGGGTCCACGCCACTC 19071

Qy   2544 CGGGGCCTTCCTGGCGACGACGACTTCCGTGCGTACCTGGGGGGACAACTACCTGCCCCAG 2603
               |||  | ||||||||||||||||||||||||||||||||    || | ||||||
Db  19072 CACCTCGGCCTCCGGGACGACGACTTCCGTGCGTCGCTGCTGCGGACCGCTCGCTGCTGCGTCGTGGCGG 19131

Qy   2604 GGCGGTCACCGACTTCGCCGCCAATCGGCCGCGATCCGCGGCGGCCGCCCGGCTGCTGCCCAG 2663
              ||| || | |||| ||||||||||    || | |||||||||||||||||||||
Db  19132 GGGATCGACGACCACCGCGTCCGCGTTCCGCCTTCGGGCCTCGGGGCGAAACCACTCCCAGCGAAACCACTCGGCC 19191

Qy   2664 ACCCCGGTCGCGGACCCGTCGGCCCTTCGGCCGGCCGCCATCGGACCGTCACGACGGGAACGTG 2723
              ||||||||  |||||| | || |||||||||||||||   | ||| |
Db  19192 ACCCGGGTGTCCGGACCCCGTCCTTCGGCGTTGGCGTGGCCGAACTGGACGAGCGCGGCCGGGTG 19251

Qy   2724 CTGCGCTTGGAGGAGAAACCCGACGTCCGGCCCACCCCGGCCAGCGAACCTGGCACTGCACGTGCGGCGTGTAC 2783
               |||||||||||||||  |  ||  ||   |||||||||||||||||||||||||||||
Db  19252 CGTCACCTGGAGGAGAAGCCGACCTGGCGCCAGCCGCCTGGCCACCTGCACACTTGGTCGCCGGCGTGTAC 19311
```

```
Qy   2784 GCCTTCAGCGCCCGGCCGTCCAGAGGCCGGTACGGGCCATCACCCCCGCCGCGGGGCGAG 2843
               |||  ||| |     ||||||||||||||||||||||     |  || ||||||||
Db  19312 CTGTTCTCCCCCGATCCAGAGGCCCATCGAGCGCCGTCCTGGCGCTGGGGGCCGAG 19371

Qy   2844 CTGGAGATCACCCACCGCCGTCAGTGATGATCGACGGCGCC 2903
               |||||||||||||||||||||||||
Db  19372 CTGGAGATCACCCACCCGCCGTGAGTGCGTCATCGACGAGGGCGAAGGACGTGCCGGTCACC 19431

Qy   2904 ACCACCACCCGGCCCTGGCCGCAGCCGCGGAGGACACCGGCCAACGCACCGACATGCTGGAGGTCAACCGT 2963
               | ||||||||||||||||||| || ||||||||||||||     ||||||||||||||
Db  19432 CTGATCAGCGAGTCACTGACTGGAGGACACCCGGCCAACGTCACCGACATGCTGGAGGTGAACCGG 19491

Qy   2964 CACGTCCTGGACGGCCGACTGTGAGGGGAGGTCACGCGCCACAGCACGCTG 3023
               |||||||||| ||| ||| ||||||||||||||||||||||||||||
Db  19492 CTGGTGCTGGAGACGGTCGAGCGTGCGGACGAGGCAGCGGAACTC 19551

Qy   3024 GTCGGCCCGGTCCGGGTGGCCGAAGGCGGCGATCGTGCGGGGTCACACGTGGTGGGGCCC 3083
               |||||||||||||  |  | |||| |||||||  ||||||||| |  ||||||||||
Db  19552 ATCCGGACGGCGTCCGGGTCGACGAGGGCCGCGTCATCCGGACGTGATCGGGCCC 19611

Qy   3084 GTGGTGATCGGCGGGGTGCCGTCGTCAGCAACTCCAGTGTCGGCCCGTACACCTCCATC 3143
               |  |||||  |  |||   |    ||||||||||||||||||||||  ||||||||
Db  19612 GTCGTGGTCGGCAGCGGCACGGATCACCACGACTCGTACGTGGGGCCGTTCACCTCCATC 19671

Qy   3144 GGGGAGGACTGCCGGTCGAGGACAGCAGCGCCCATCGAG---TACTCCGTCCTCGTCGCGGGC 3200
               |||  ||  ||  | ||||||  | |||| |       | |||
Db  19672 GCCGAGGGCTGCCTGGTGAGGACCACCAGCCGAGCTGGAGTTCTCCATTCGTGCTGCGCAAC 19731

Qy   3201 GCCCAGGTCGAGGGGGCGTCCGGATCGAGGGGTCCTCATCGGGCGGCGCCGGCCGTCGTC 3260
               |||  ||||  | |  | |||||||||||||||||||
Db  19732 GCCTCCATCTCCGGAGTGCCGGCGTCCCTCATCGGGCGCACGTCCGGGTG 19791

Qy   3261 GGGCCGGCCCCGTCTCCCGCAGCGTCACCGACTGGTGATCGGCGACCACAGCAAGGTG 3320
               |||||||||||||||||||||||||||||||
Db  19792 ACCGGCCGGCCCCGCTACCCGCGTACCCGCTGGTCCTGGCCGACCAGCACCGGCC 19851

Qy   3321 TATCTCACCCCATGACCACGACC 3343
              ||||||||||||
Db  19852 CAGAATCTCCTCCTGACCCCGATCGACCACGCCGAGAGACGGAGAAGGACGGCTCCGCCTACGTCGCGCCG 19911

Qy   3344 ------ATCCTCGTCACCGGGGCCCGGGGCGGCCCCCGGGGCCCGCCGCCAACAAGCCC 3394
                    | ||||||||||||||||||||||||||||||||||||
Db  19912 GCCACACGTCTGCTCGTCCTGCCGGCGCCCGGCGATTCATCGGCTCCACTACGTCCGTCCTCGACAAACTCACCTAC 19971

Qy   3395 CTCCTGTCGCGGGGGCCCCGGGGCCGGGCGTCCGGTCCGGGAGGCGGGCGTCGTC 3454
               | || |||||  |     | ||||||||||||||||||||||||
Db  19972 CTCCTCGGTCCGGACGGGCCGCCGAGCGCTCGTCGACGCGCTGTCGACGGGCCCTGAGCTAC 20031

Qy   3455 GCCGGCAGCCTCGCCGGGAGATCCGCGGCCGTTGCTGACCATCCCCGGCTCACCTTCGTCCAG 3514
               |||||| | || |||||||||||||||||||||     ||||||||
Db  20032 GCCGGCAACCTCGCCAACCGCGAACCTCGCCAACCCCGCCGCGGCGGGACGACATCGTG 20091

Qy   3515 GGCGAGCTGTGGCGACACCGTGGCCGACACGCTCGACACGGCGACACCGTG 3574
               ||||| |||||||||||||||||||    ||| || ||||||||||
Db  20092 GGCGACATCTGGCGGCCGACCGGACGCGGTCCGGGTCATGGCCG---GCCAGGACCAGTCGTG 20148

Qy   3575 CACTTCGGCGTGCCGGGGGGGACAACTACCTGCCCCAG 3634
              ||| || ||||||||||||||||  |  | ||  ||  ||
Db  20149 CACCTCGGCGGCGGGCCGAGTCGATGCGCGCTCGCGATGCCGCGCTCGAGCCGCGTGTCGCGG 20208

Qy   3635 ACCAACGTGCTCGGGCACCCAGTTCGGCGCGGGGACAATTTCGTGGTGGCTGCTC 3694
               ||||||||||||||||||  ||| |||||||||||||||||
Db  20209 ACCAAGCGCGGCCGGCGGACAGCCGCGACACTCCTCGACACTGCTGCTC 20268

Qy   3695 TTCGTGCACGTTCACCGACGAGGTACGGCCTCCTCCCGGGAG 3754
               ||  |||| ||||||||||||||||||||||||||||||
Db  20269 TTCGTGCAGGTGTCCAGCGAGGTCTACGGCGTCTACGGCGGGGTCGTTGGACGGGAG 20328

Qy   3755 AGCGACCCCCTGCTTCCGACCCGTCCGCCGCTCGCTCATCGGCGTGTCATG 3814
               | |||||||||||||||||||||||  ||     |||||||||||||
Db  20329 GACGAGGCCGTGCGCCGCCGCCGACCTGCCAACAGCCCCTATGCCGCGGCCGTCTTGCGCGGCGACCTGCTC 20388
```

```
Qy    961 GGGAAGCGGGACTCCAGGCGCGCAGCCGCTCCTGGATGAGCTCGCCGAGGACGGCGCACG 1020
              ||||  ||||   |||||   ||     ||||  ||| |  |||||| |||| || ||
Db  46149 GCGAACTGGATGCCGCGGTGCCGTGCGAGTCGTCGAGTTGATCGATCCCGGTAGTGGGAC 46090

Qy   1021 CGGTCGATCAGCCGGGGGTGCGCGTCGCGCTGCTCCTGACCAGCCTCGTGTCC 1080
              ||||| ||   ||   ||||   |  |  ||   ||| || ||||   |||
Db  46089 ACGTCGATGCCGCACAGTTCGGGGCACGCACGCGGCCGGGTGGCCGCCTCGCCG 46030

Qy   1081 AGTGGGTTGCTCCGGCGACGGGCGGCACGTGCGCGACGTGGCTGCGCACGTCGAC 1140
              ||| ||  | || ||||| ||| | | |||||||||   |||   || |||
Db  46029 GCAGCGGTGGCGAGGCTGGCACCGGCGGTCACGGGACGTGGCTGCACGTGGAC 45970

Qy   1141 GCTTCCGCGCGTCGAACTCGTCCGCGTGCCGTCCGGCCAGCAGCACGCG---GAAGGGGAATCCGCTCGGCGGTGCCCTTGA 1196
              |||   |||  ||||  || | ||||     || ||| ||  || ||      |||  ||  ||| || | ||||| ||
Db  45969 CGGGTCACGGACGGCGGTATCGGACGGCGTGCGGGCCCGCCGGGGCCGGTCCATCTTTGAAGAAATCCGC 45910

Qy   1197 CAGCATCGCCAGGTCCGGCTCACCTGGAGCTGCTCGCACGTGTGC 1256
              |||  ||| ||||  | ||||| |  ||| || | | ||| |||
Db  45909 CTAGAAGGCAAGAGTGACCGAAAATGCCTGTGTCGTCCGACAGCAGCAGCCGGTGGGGCCG 45850

Qy   1257 CCCGCGCGGTGAGGACCTCGTCGGCGGCCATTCCAGCTGTCCGGCGCCTAC 1316
              ||   || ||||  |||| | ||||  ||     ||| |||||| | |||
Db  45849 CCCGAACCCGGCACGTCGTGGTCGTCGGCCACGCGCCGGGCCTGGTGGTGCC 45790

Qy   1317 GGCGATCCGCTCCCAGTAGCCGGGAACTCGGTTACCGCGCGCCGCCGACTTCGCG 1376
              ||| ||||| |||  |||  | |||| | |    ||||||| || ||  ||||
Db  45789 GCCCTGACGCGGCGCACCGGACCTGGGACACGTCCCCGCTC------ 45739

Qy   1377 TTCGAAGACCAGGGCCGGCACGTGTACGAGCGGCGGCGACGGTGGCGGCTCCCCTG 1436
              |||  | |||  | || ||| ||| | | | |||| ||||| | ||||| | ||
Db  45738 ------CGATCGACGCCGGTCCTGTTCGACTTCTCCGGCACGCTTCCGGCGGTGAACTGCC 45684

Qy   1437 GCACCGCACGTCGCAGAGGGGTACTCCAGGGCCACACCGGGC 1496
              | || | ||| | || ||||  || | |||||||| || ||
Db  45683 GCAGCGTTGGCTGCGGGGGCGCGGTCCAGCTGCGGACACCGGGGACGTACC 45624

Qy   1497 TGTAGCCAGCACGCTGTCGCCGGTGAAGGCTCGTGCGTCCCAGTGGACCAGCAT 1556
              |  |||||| || | |||| ||||   | || |    ||||  |||  |  ||
Db  45623 GGAGGCCCTGACCGGTTGCGTCCGTGGCGGCCGGTCC--CGGCGGGCCGGAGCCGGAC 45566

Qy   1557 CCGGCGGCGGGTCCGCTCCTGACGACCCGCAGGGCGGCCTCGACCTCCGGCGGTGA 1616
              ||||| | || |||| |  |||||  |||||||| ||||  | |  ||| |||||
Db  45565 GCCGTCGACCGGGTCGCCGAGTGTACGCCCAGGGCGGCCTCGACTCTCGCGAACCGAC 45506

Qy   1617 CGGCGGGCGGGTCCGCTCGCCTGACGACCCGCAGGGCGGCCTCGACCACCTCCGGCGGTGA 1676
              |||||||||||  || ||                ||||| |||||| | || ||| |||
Db  45505 CGCACGC------CGTACGGTGGGCTGCTCGCCTGACGACCTGGCCCATCCCGGCCTC 45452

Qy   1677 GAAGAAGGCGTAGTGTCGAGCTGTTCGGGCAGCTGTCCAGCAGTTCCAGCAGGCC 1736
              |||||||||||| | ||    ||||| ||  || ||  |||  |   || |||||
Db  45451 GCCGAGGGCCTCTACGACGTCGTTCCCACGACGCTGGCTGCCCTACCGCGACGCGCC 45392

Qy   1737 GGCGCCGGTCCGGGTGCGGTCGTCGTGTGGCGGCACCAGGCGGCGGCGTCGGGTGGTGAG 1796
              |||  ||  | ||||| |||  |||||    ||| |||| | || ||  | || |||
Db  45391 GCCCCGGTCC----TGGGGAGTTGCGGCCGGCGGTCGGGCACCGGCGTGGGCACCGGTCAGCAA 45336

Qy   1797 TGCCTCGACGACCTCCGGGTGCCCGGTGGCCCAGTGACTGGTCCGGCCGGA 1856
              |||  ||||| ||| || || ||| ||||| || ||||| ||| | ||||
Db  45335 CATCGCCGTCGGACCTGGACCTGCGGCCCACCTTCGGCGGCGGGCGTGGTGGACGC 45276

Qy   1857 GTCGAGGTACTGTTGCCGTCAGGTCGTCAGTGTGCCTGCGAAGAC 1916
              ||| || ||||| |||| | |  ||||| || ||  ||| ||||
Db  45275 CTACGTCTTCTCCCACGAGTGGGGCGCGTTCCTGCGCCGATCTTCCCGCACCGC 45216

Qy   1917 CCGGGTCCGTCGACGCCGGTTCCTCGGAGCGGCCTCCGGCGGTGGCCGGTCG 1976
              ||| ||  || ||  ||||   || |||| ||     |||| || ||| ||
Db  45215 CTGCGACGAGCTGAAGGTGCCCCGCA-GCGGGCGCTGATGTGGGCGACACCCTGGCCG 45157

Qy   1977 CAGTGCTGCTGCTGCGCTAAGCCTGTCATCGCGCTGACG 2036
              |||| ||| |  | ||     |  ||| | ||| |||||
Db  45156 ACGGTGGGGCGGGTGTACGCCGGCGCTCGCAGACCCTGCTGCTGCGCCCCCGGACA 45097

Qy   2037 CGAATCGCCGGCGAACTGCCGTTGTGGCGGCACCACGGTTGGGCGCGCGTCGGCTGAGTCAA 2096
```

```
Qy   3815 GCGCTCGCCGCCACCACGCACCGGCCTGGACGTCGGGGTGACCCGGCTGTTCGAACAAC 3874
              ||||||||||    || ||| |||    ||||| ||   |||| ||||||| ||||||
Db  20389 GCGCTGGCCGCCATGACGACGTCAGCCGGGCCTGCCGCGTCGCCGCATCACCCGCTGCTCCAACAAC 20448

Qy   3875 TTCGGCCCCCACCAGAGCTCCGAGAAGCTCATACCGCCGCTTCCTGACCAGCCTCCTGTCC 3934
              |      |||||   ||||   ||||||| |   ||||| ||   |||||||||| ||
Db  20449 TACGGCCCCTACCAGTTCCGGGAGAAGCTCGTTCCCGACGGTCGTCACCGGTCCGTGCTGGAG 20508

Qy   3935 GGCGGCACCGGTTCCCCTCTACGGGCGACGGCGGGCACGTGCGCGACGTGGCTGCGCACGTCGAC 3994
              ||| || ||  | |  || || ||||||||||||    |||||||||||| |||| |||||||
Db  20509 GGCCCAAGGTGCGCCGTTGTACGGGCGCACGGGCGGTCACGGGACGTGGCTGCGCACGTGGAC 20568

Qy   3995 GACCACGTCAGGGGCCGTCGAACTCGTCCGGCCCGTGCCCGGCCCGGCGGAGATCTACAAC 4054
              |||||| | ||  |||||| || | ||  |||| | ||  |||||  | ||| | |||||
Db  20569 GACCACGTGCGCGGTATCGAGGCGGTGCGCGGCGGTGCGGGGCCCGGCGCCTCTCACAAC 20628

Qy   4055 ATCGGGGGCGGCACCTGCGCCCACCTGAGCTGCCGGTTGCTCGCACGTGTGC 4114
              ||||||||||||||| |  ||  ||||||| || ||  ||||  ||| |||
Db  20629 ATCGGAGGCGGCACCTGCGTTGGCCAACCTCGGGACCTGCTGCTGAAGGCGTGC 20688

Qy   4115 GGCGGCGGGCCCGGAGCGCATGCGTCCACGTCGAGAACCGCAAGGGCACGGCCGCCGCCTAC 4174
              |||  ||||| ||| ||| || | | |||| ||| |||  |||||||||| || ||||
Db  20689 GGCGGGCTGGACGCACGGTGGAGCAGCGTCGGGACACCGGGCCAGGGCCAGGACCTGGTGCC 20748

Qy   4175 GGCGTCGACCACACAGCAAGATCACCCGGCGAACTCGGTTACCGCGCGCCGCCGACTTCGCG 4234
              |||||| || ||| |   | ||  | |||| ||| |||    ||||||  || | ||| |
Db  20749 TCGGTCGACGCGGCGGGATCCGCGGCGTGGACCACGTGCCCCGCGGAACCGGACCTGAGC 20808

Qy   4235 ACCGCCTGCGGACCGCCGAGTGTACGAGCGGCGGCGACGGTGGCGTGGCGCTCCCCTG 4294
              ||||| |||| | ||||||||| | | |  ||||||||| ||| || | |||||| ||
Db  20809 ACCGGACCTGGGCCGGCGGACAACCGCGGTGGCTGGTACCGCGCTGGGCGGGCCCCCTG 20868

Qy   4295 CTCGCCCGGACATGACGTCGGGGCCCGGACCGTGCCCGACCCGGCGCACACCGCCC 4354
              || |  | || |||  || |||  | |||   |  || || | ||||||||| |||
Db  20869 TGCGGCCGGCGACCCCGCGCCCGATCCCCGGGACGTACCCGCCCGGCGGCACCGGGC 20928

Qy   4355 CCCGGGCCGGTCCGGGTCCGCTCAGCGTCCGCAGCGCCCCCCGCCCCCGGGGC 4414
              ||  |||| | |||  |||  ||| || |  |||   ||| | || ||||||
Db  20929 GGCGCGGGCCGTCGACCTCGGCCGGCCGCA------CAGCGCCGTCGGCCGCGGGGCCACGCGC 20982

Qy   4415 CGGCGGCGGTGGACCCCGG 4434
              |||| ||| ||| ||||||
Db  20983 CGGCGGGCCGCCGCCAGGCACCG 21002
```

```
RESULT 8
AX195929/c
LOCUS       AX195929        109519 bp    DNA        linear   PAT 28-AUG-2001
DEFINITION  Sequence 1 from Patent WO0151639.
ACCESSION   AX195929
VERSION     AX195929.1  GI:15386161
KEYWORDS
SOURCE      Micromonospora carbonacea.
ORGANISM    Micromonospora carbonacea
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Micromonosporineae; Micromonosporaceae;
            Micromonospora.
REFERENCE   1  (bases 1 to 109519)
AUTHORS     Hosted,T.J., Horan,A.C. and Wang,T.X.
TITLE       Everinomicin biosynthetic genes
JOURNAL     Patent: WO 0151639-A 1 19-JUL-2001;
            Schering Corporation (US)
FEATURES    Location/Qualifiers
source      1..109519
            /organism="Micromonospora carbonacea"
            /db_xref="taxon:47853"
BASE COUNT  15101 a  39910 c  39406 g  15102 t
ORIGIN
```

Query Match   12.8%;  Score 877.8;  DB 6;  Length 109519;
Best Local Similarity  56.5%;  Pred. No. 7.4e-59;
Matches 1892;  Conservative   0;  Mismatches 1402;  Indels  53;  Gaps  12;

Left column:

```
Db  45096  |  |||   |   |   ||    |         |  |      ||        ||
           CCGTGCACGG-----CCTGGCGCCGTGCTCGAACTGGTCGGCGGTGTTCAGAAAGTTGTTGC--GAG  45041

Qy  2097   ACACTTGAACACACACCGCTGCCAGAGTTGCCGGTTGTTCAGAAAGTTGTTGC--GAG  2154
           ||  ||      |  |     ||  |  ||      |   ||
Db  45040  CCGTTTGCGGAAATGGCCGACGCGCATAGAGATCCACCGATCACCACGATCCGAC  44981

Qy  2155   CGGCCCCGGCACTCTGGTTGAGTCGACGTGTTACGGCGCCACCACGCCTCACGTTCGAG  2214
           ||    ||  |||||||||
Db  44980  TAGGGAAATCACGAGTCAGAGTGATCCTTCGGTACGTTCACGAAGAGCCCAGTCACCGAC  44921

Qy  2215   GAGGGACCTGTGAGAAACAAGCCCAGACCGCCCGCTCGAACGGCCG-AGGTGAAG  2273
           |||                           ||||||     |  ||
Db  44920  TGGAGTGCTCGTGAACCAGCGACCGTCTCGGTCCCATCCTTTCGAGGGCGTTCGTGAAG  44861

Qy  2274   GCCCTGGTCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACCGCCGCC  2333
           || |||||||   ||    ||   |   ||||||||||    ||||
Db  44860  GCCCTGGTCCTGGCGGGAGGGGGTGGGATTGGAGATTGCGTCCGATCACCCACACGTCCGCC  44801

Qy  2334   AAGCAGCTGCTCCCATCGCCAACAAGCCGTGCTCTTTCTACGGCGTGGAGTCCCTCCC  2393
           |||   ||   ||||  ||  |  |||||     ||    |||||||
Db  44800  AAACAACTTCATTCCCGTGCCAACAAGCCGGTGCTCATCTCCGGCTCGTTTCTACGGCGTATTCGT  44741

Qy  2394   GCGCGGGTGTCCGGGAGGCCGGCGTCGTCGTGGCGCGCTACGGCGTACGGCCGGAGATCCGCGAA  2453
           | ||||||   |||||         ||||      |    ||||||||
Db  44740  GAGGGCGGTATCGAGGAGGTGGGCATCGTCGTGCCGCAGCAGCCGCACCCGCGCCCCGCGCCCCG  44681

Qy  2454   CTCACCGGCGACGCCGTTCGGGTTACGGCATCACTACCTCCACCAGCCCGCCCG  2513
           | |||||||||   |||  |||||  ||    ||||||||  |||
Db  44680  GTGGTCGGCGGATGTTCCCAGTTCGGGCTCCAGGTGACCTGTCCGCCACGCTCCC  44621

Qy  2514   CTCGGTCTCGGCCACGCGCCACGCCGTCGGCGCATCGCCGGCGCTTCTGGGGCGACGACTTCTG  2573
           |||||||||   ||||     ||  ||  |||   |     |||
Db  44620  CGGGGCTGGCGACGCGCCACGCCGTCCATCTCCCGGCGACTACCTCCGGCGACGACTTCGTC  44561

Qy  2574   CTGTACCTGGGGACAACTACCTGCCCCAGGCGGCTCACCGACTTCGCCCCAATCGGCC  2633
           |||||||||       ||| ||         |   ||||      |||
Db  44560  ATGTACCTCGGCGACAACTTCATCGTCGTGCCGCGCGGATCAAGGACATCGTCGGCGCTTCGG  44501

Qy  2634   GCCGATCCCGGCGGCGGCCCGTCTGCTGTCACCCGGTCGGCGCGGACCCGCTTCGGC  2693
           | |||||       ||   ||    ||||||||||||     ||||   |
Db  44500  CAGGAGCGACGCGCAGCGCAGATCATGCTGACCTGCGGCGTCGCCGCACCGCCTTCGGC  44441

Qy  2694   GTCGGCGGAGGTCGACGCGGACGGAAGCGACGTGCTGCGCTTGGAGAGAAACCGACGTCCCG  2753
           |||||    |   ||||||||||||         |  |||||||
Db  44440  GTCGGCGAGATGGGCGCGGACGCGGGTGATCGGGTTCAGCGCCGGGCGGGCGAGCCACCCC  44381

Qy  2754   CGCAGCTCGCTCGCCGCTCATGGCCGTGTACGCCTTCACGGAGACTCCCACGAGGCGGTA  2813
           ||      ||    ||  ||       |||||  ||       ||||
Db  44380  AAGAGCGACCTGGCCTGGGCCCTCTTCAGCGTCTTCAGCCCGGTGGTCCACGATGCCGTG  44321

Qy  2814   CGGGGAAGGTCGACGCGCACGCACCGCTGGTCGGCGATCGAGATCACCCCACGCCGGATG  2873
           |     |  |||      ||  ||||  |||       |||||  |  |||
Db  44320  GCCGAGCTGAAGCCGGCCGCTCGCCGCCAACGACCCGAGGCCGCAGCACGCCGGATG  44261

Qy  2874   ATCGACCGGGCGCGCACGGGGTCGACTCCACCATGATCACGGGCTACTGGAAGGACACCGCGG  2933
           ||||||||||||   |||      ||||||    ||||
Db  44260  ATCGACCGGGCGCGCACGGGTCGACTCCACCATGATCACGGGCTACTGGAAGGACACCGGC  44201

Qy  2934   AGCGCGGAGGACAGTCTGAGGGTCAACGGCTCGGCGGGGACTGGAGGCGGGGCCGCCATC  2993
           ||  |||||||||||   ||    ||||
Db  44200  AACTTCCGCGACGATGCTGGAGATGAACCGGTCGTTCGCTTCGCCAGCGTCGGTCCTTTGGTC  44141

Qy  2994   GAGGGGAAGGTCGACGCGCACGGGTCGGGCGCACCGCTTCGGTTCGCGAAGGCGCG  3053
           |||||||  |||  |  ||  ||||||
Db  44140  CTGAGCTGTCCAACATGGAAGCTGATCCCGGTCGTGGTACGGCGCGCCTTGGGACGCCGGC  44081

Qy  3054   ATCGTCGGGGAGGGTCCACCGTGGCGCGGTCGATGCGCGGTCGGCGTCGTCGAGCAGG  3113
           ||||       |        ||||||||| |||||   ||| ||||
Db  44080  GACGTCAGCGGCGCTCAGCGTGTCCCGGTCGGTGGTGTGCGTCGATCATCCCGC  44021

Qy  3114   AACTCCAGTGTCGGCGATCTCGGGGAGGAGGACTGCCCGGTCGAGGACAGCCGCC  3173
           ||||||   |||||||
```

Right column:

```
Db  44020  AACTCCTACCTCGCCCGCTACACCTCCGATCGACTGCGACCTGCTGCCGGACTGCTGCGACCGGAG  43961

Qy  3174   ATCGAGTACTCCGCCGCTCGGTCGCCGCGCGCCCAGGTCGAGGGGGCGTCGAGGGGGCGTCGAGGGCG  3233
           ||||| ||||  ||||  ||||           ||||||||
Db  43960  ATCCAGCACTCCATCGTCTGCGCCGCGCGGGCGATCGGCCGCATCGAGGCTCACCGA  43901

Qy  3234   TCCCTCATCGGCCGCGCCGTCTGCGGCGGCGGGAGGCGGGTGCGGGTCCCCGGGGCTGCCCAAGAGGCCACCGG  3293
           ||  | |||||||  ||||  ||    |  | |||||||  |||  ||   |||||
Db  43900  TCGATGATCGGCCGGTCGGATCACAGCGGGTCGGCGTCCCGGTGCCAAGAGGCCACCGG  43841

Qy  3294   CTGTGATCGGCGACCACAGCAAGGTGTATCTCACCCATGACCGATGACCATCCTCGTCA  3353
           || |   ||||  | |||||     ||  |||||||   ||||||
Db  43840  TTCGTCCTCGGCGATCACAGCGAGTCCGGCGTGGGGGGTCTGATGCGTCGCGTCCTGGTCA  43781

Qy  3354   CCGGGGAGCGGCGGCTTCATTCGCTCCGGCCTACGTCGTCCGCGCTCCTGTCGCCCGGGGCCCC  3413
           |||||   ||   |  ||  ||||||||      ||   ||||
Db  43780  CCGGGGTGCCGGTTCATCGGCTCCAGTCGCTCCGGCGACCGACGCCACCGGGTACTACC  43721

Qy  3414   CCGGC------GGCGTCGGTGACCGTCGCTGACTGACTGAGCGTCCGCCGGC  3467
           ||| |
Db  43720  CCGACACGGCGCAGGCCCGGTGCTGGACAAGCTGACGTCGGCGGCCACCTCG  43661

Qy  3468   CCCGCCCGCACGCGGTGCCTGACGTCCAGGACC----GGATCACCTTCGTCCGCCGGGAGGCCCG  3527
           |||   ||  |||||||||||| ||||||||||    |  ||||||||||
Db  43660  CCAACCTCGAACCGTCCAGGAC------GGATCACCTTCGTCCGCCGGGAGGCCCG  43607

Qy  3528   ACACCGGCGCTCGTGACACCTGGCGCGCGGCACGACCATCGTCGCTCGGGCCG  3587
           |||| ||                  |||||||  |||
Db  43606  ACACGGCGGTCTACGGCGCGGTGCTGCCGGAGTGCGGCGGTGCTCAACTTCGCCCCG  43547

Qy  3588   AGTCGCACGTCGACCGGCTCCATCACCGACAGCGGTGCCTTCACCCGCCAACGTGCTGG  3647
           ||||||||||||   |||||  ||   |||||||
Db  43546  AGTCCGCACGTCCCATCGCCACCGCCGGTTCGTCGCCACCAACGTGCAGG  43487

Qy  3648   GCACCCAGTCCTCGACGGCACGGTGTGCGCCACCTTCGTCTGCGCACGTCT  3707
           |||||        |   ||||||||||
Db  43486  GCGTCCAGACGCTCATGCAGGCGTGTCTCGACGGCGCGGGTGCCGGTCCAGGTCT  43427

Qy  3708   CCACCGACGAGGTGTACGGCTCCCTCCCGGGCGCGGGGAGAGCCACCCCTGC  3767
           ||||||   | ||||||| | |||||
Db  43426  CCACCGACGAGGTCTACGGCTCGACGAGGCCGGAAGAGGACCACCCCCCTGG  43367

Qy  3768   TTCCGACCTCGCCGTCGACGCCGTCGGCGACCTCATGGCGCTCGCCCACC  3827
           |||||||||||| || |  |||||||||||||
Db  43366  CGCCGAACCTCGTACGGCGTACGCCGTCGAATCGCGCCGATCGCGCCTACG  43307

Qy  3828   ACCGGACCACCGGGCGGGCCGGAGGTCGGGTGACCCGTCCAGGCGGCGGTA  3887
           |||||||  ||||||     ||||||  |||
Db  43306  CCCGCACCCAGCGGGCTCGCCCGTCGCCGCCAACAACTACGGCGCCTACC  43247

Qy  3888   AGCATCCGAGAAGCTCATACCGGCCTTCCTGTCCGGCGCACCGGTTC  3947
           ||| ||  |||||||  |  ||    || |  |||||||
Db  43246  AGTTTCCGGAGAAGCTGATCCCGCGTGTTGACCAGCGGCGACGGGTGC  43187

Qy  3948   CCCTTACGGCGACGTCGAGAAGCGCGCCACGACCGCGGG  4007
           || ||| |||  |||||||||    ||||||||||
Db  43186  CGCTGTACGGCGACGGGCGAACGTTCGGACGCTGACCACTGCCGGG  43127

Qy  4008   CCGTGCAACTCGTCGCGTGTCGGCGGCGGCCAGGCCGCA  4067
           ||  |    |||  |||||||| |    ||| |||||||
Db  43126  GCATCCAGACCGTCGTTCGACCGGCGCGTGTCTACCACATCCCGCCGCGGCACGG  43067

Qy  4068   CCTCGCTGCCCAACCTGGAGCTCGGCGCGCTGGTGCTCACGACCACCGGTTC  4127
           |||||||  |||||  ||   |||||||||  ||||
Db  43066  CTGAGCTGTCCAACATGAGCTGATCCCGGTCGTGGTACGGCGCGCCTTGGGTGGG  43007

Qy  4128   AGGGACATCGTCCACGTCGAGAAACGTCGCGCAGCCGCGGCGTACGGCGTGACCACA  4187
           ||||     |||   ||||||||||      ||
Db  43006  ACCGGGTCGAGCGGTCGACGGGTGCCGACACCGCAAGGCCCAGAACCGCGACTGCCCC  42947

Qy  4188   GCAAGATCACCGGCGTCGGTTACCCGCCCGG  4247
           |||  |||||||       ||    ||||
Db  42946  CGAAACTGCCGGCG---CTCCGCGTACGGCCCGCCCAGTGTCGGCCTCGCCGACGCTGCCCG  42890
```

```
Qy   4248 ACACCCGGAAGTGGTACGAGCGGCACGAGGACTGGTGGCGTCCCCTG 4294
           | ||  |  |||||||| ||| | |||  |||||||||||  ||| |||
Db  42889 AGACGATCGACTGGTACCGGACGCACCGGGCACTGGTGGGAGCGCGCT 42843
```

RESULT 9
STMDNRLM

```
LOCUS       STMDNRLM                2682 bp    DNA     linear   BCT 28-MAR-1997
DEFINITION  Streptomyces peucetius glucose-1-phosphate thymidylyltransferase
            (dnrL) gene, complete cds and truncated dnrM gene, complete cds.
ACCESSION   L47163
VERSION     L47163.1  GI:975620
KEYWORDS    daunosamine biosynthesis; dnrL gene; dnrM gene; glucose-1-phosphate
            thymidylyltransferase; mutation; truncated form.
SOURCE      Streptomyces peucetius
  ORGANISM  Streptomyces peucetius
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 2682)
  AUTHORS   Gallo,M.A., Ward,J. and Hutchinson,C.R.
  TITLE     The dnrM gene in Streptomyces peucetius contains a naturally
            occurring frameshift mutation that is suppressed by another locus
            outside of the daunorubicin-production gene cluster
  JOURNAL   Microbiology 142 (Pt 2), 269-275 (1996)
  MEDLINE   97086506
   PUBMED   8932700
FEATURES             Location/Qualifiers
     source          1..2682
                     /organism="Streptomyces peucetius"
                     /db_xref="taxon:1950"
                     /tissue_lib="ATCC 29050"
     gene            655..1710
                     /gene="dnrL"
     CDS             655..1710
                     /gene="dnrL"
                     /function="involved in daunosamine biosynthesis in the
                     daunorubicin pathway"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-phosphate thymidylyltransferase"
                     /protein_id="AAB50923.1"
                     /db_xref="GI:975621"
                     /translation="MKAIVLSGGSGTRLRPFTHTSPKQLVPVANKPVLYYVLEDIAQA
                     SITEVGIVVGETSNEIRKAVGSGDRFGLRVTYLPQEAPLGLAHAVLIARDYLGEEDFV
                     MYLGDNFVVGGIAGNSSTFRAERPDAQILLTRVSDPSSFGVAEIGCDGRVVALEEKPR
                     HPRSDLAVVGIYLFTPVVHEAVRALTPSRRGELEITDALQWILLDGPYDVRYTTISGYW
                     KDTGNVADMLEVNRAVLDGIEPGMEQADAASELVGRVRIEAGAQIRASRIVARRHRA
                     GRVTDRTSALHVDRGDCSIETARSSLHHAGRLPAHGTRRVQHSLLGRNVTVAPARVP
                     AGSRLILGDDSRVEISS"
     gene            1707..1892
                     /gene="dnrM"
     CDS             1707..1892
                     /gene="dnrM"
                     /function="mutant gene with frameshift at nt 1707 that, if
                     corrected, would produce a thymidine diphospho glucose 4,
                     6-dehydratase for daunosamine biosynthesis in the
                     daunorubicin pathway."
                     /note="truncated"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB50924.1"
                     /db_xref="GI:975622"
                     /translation="MTMKILVTGGAGFIGSHVRALLGPRGDGSVTAGQPDLCRQS
                     GQPRSRPRRRAVRLRPR"
BASE COUNT      438 a    939 c    873 g    432 t
ORIGIN
```

Query Match 12.8%;  Score 875.4;  DB 1;  Length 2682;
Best Local Similarity 66.4%;  Pred. No. 4.4e-58;
Matches 1368;  Mismatches 0;  Conservative 0;  Indels 17;  Gaps 7;

```
Qy  2241 GACCGACACCGCTCCCGGAGGCGGCGAGGTGAAGGCCCTGGTCCTGGCAGTGGAACCGGCC 2300
          ||     | || |||||| || |  |  ||||| |||| |||||  |||||||||||||||
Db   628 GAGCGGCGGCCGCCGCCTAGGCTCGGCGGCCATGAAGGCTCTGCTCCTGCCGGGGGTCGGG 687

Qy  2301 AGCAGACTGAGGCGCCTTCACCCAGCACTGCTCCCATCGCCAACAAG 2360
          | | || | || ||||||||||| || | || | ||| |||||||||
Db   688 ACGCGGCCTGCGGCGCCCTTCACCACACACGTCTCCAAGCAGTTGGTGCCGGTAGCCAACAAG 747

Qy  2361 CCCGTGCTCTTCTACGGCGTGGAGTCCCTGCGGCGGGTGTCGGGAGGCCGCCGTC 2420
          || ||| | ||| |||||  || ||| |   |||  |||| ||| | ||| |||||
Db   748 CCGTCCCTCTACTACGTCCTCGAGGACATCGCGCAGGCGTCGATCACCGAGGTCGGCATC 807

Qy  2421 GTCGTGGGCGTACGGCGCGGGATCCGGGATCAGCCCCGGTCGGCTCGGCACGCGGTGCGCATC 2480
          ||||| ||| |  |   ||||  |||||  || ||| ||||| |||||||  |   |   |
Db   808 GTCGTTGGGAGACGTCCAACGACGATCCGCAAGGGCGTCGGAGGCGGCGCAGAAGATTCGGG 867

Qy  2481 TTACGGCATCACTACCTCACCAGCCCGCTCGGTCGGCTCGCGCACGCGGTGCGCATC 2540
          |||||  |  ||| |||| ||  |  |||  ||| |||||  || |||||||||||||
Db   868 CTCCGCGTCACCTACCTTCCGCAGGAGGCCCCGGGCTGGTCACGCGTGCTCCTCATC 927

Qy  2541 GCCCGCGGCTTCGTGGGCGACGACACTTCGTGTACCTGGGGACAACTACCTGCCC 2600
          ||||  || | || ||||  | |   ||  | |||||| || ||  || |||||||
Db   928 GCCCGCGACTATCTCGGCGAGGAGGACTTCGTCATGTATCTCGGGGACAACTTCGTGGTC 987

Qy  2601 CAGGGCGTCACCGACTTCGCCCGCGAACTTGTCGACGTTCGGCGGCGCGGCCCCGATCGCGC 2660
          || || ||| || ||  | || || |  |  |  ||  |  ||| |||||  |   ||| |||
Db   988 GGCGGCATCGCAGGGAACTTCGTCGCGTCCTCGGGCGCCCCCTGCGATGCCCAGATCCTG 1047

Qy  2661 CTCACCCCGTCGGCGGACCCGTCCCGCTTCGGCGTCGACGGTCGCGACGCGGGAAC 2720
          ||||| | ||  | | ||| |||| ||| ||| |||  | | || |   |||||||
Db  1048 CTCACCCGTGTCTCCGACCCGTCCAGCTTCGGCGTGGCCGAGATCGGCTGTGACGGGCCG 1107

Qy  2721 GTGCTGCGCTTGGAGGAGAAACCCGACGTCCGCGAGCTCGCTCATCGGCGTG 2780
          || | | || ||||||||||| || | |  ||| || | || |||| |||||
Db  1108 GTCGTCGCCCTGGAGGAGAAGCCGAGGCACCCCCGCAGTGATCTCGCCGTGGTCGGGATC 1167

Qy  2781 TACGCCTTCAGCCGCGGGTCCAGGAGGGTCCATCACCCCTCCGGCCGGC 2840
          |||  | | ||| |||  ||||   |||||| | || ||||||  |||||
Db  1168 TACTTGTTCACCCCGGTCGTGCACGAGGCGGTGCGCGCCCTCACGCCGTCGCGGCGGGGC 1227

Qy  2841 GAGCTGGAGATCACCCACCCAGCTGCAGTGGATGATGACCGACGGGGCC 2900
          ||||| ||||  | ||  | ||| |  |||| | || | ||  ||||| |
Db  1228 GAGCTGGAAATCACGGACGCGCTCCAGTGGATCCTCCTCGACGGTCCGTAC 1287

Qy  2901 GAGACCACCACCCGCGCCCTGCGCGACACCGCACGCGCGAGGACGATGTCAAC 2960
          |  | || || | || || | ||  | ||  ||| ||| | ||| ||| |||
Db  1288 ACCACCATCTCCGGTTATTGGAAGGACACGGGAAACGTCGCCGACATGCTGGAGGTGAAC 1347

Qy  2961 CGTCACGTCCTGACGGACTGGAGGCGACGAGCCGGCCAGGGGAAGTGCACGCACG 3020
          ||| |  |  | |  |||  |||  ||  |  ||| |||  ||| ||| ||| |
Db  1348 CGTGCCGTGCTCGACGGGATCGAGCCGGGCATGGAGCAGGCCGACGCCGCCAGCGAA 1407

Qy  3021 CTGGTCGCCGGGGGTCCGGGTCCGGGCGGCCAAGGCCGGCGGGTCACGCTGGTGGGC 3080
          || | ||  || ||||||| |   ||| |  |  ||| ||| |   ||||  | |||
Db  1408 CTGGTGGGCGGCGGTGCCGGATTCGAGGGCGGCCAGATCCGGGCGTCGCGCATCGTGGCC 1467

Qy  3081 CCGGTGGTGATCGGCGCCGGTCGCTCGTCAGCAACTCCAGTGTCGCCCGTACACCTCC 3140
          ||  | || ||  | ||| || |  |||  |  ||  |||| | ||| || |||| |
Db  1468 C------------GCCGTCATCGGGCCGCCGGGCGCGTCAGTCGACCGGATCGTGACCGTCTCGGGCTCGGCCGATCTCGGGCCCTACACGTCG 1519

Qy  3141 ATCGGGGAGGACTGCCGGTCGGGATCGAGGACTACCGATACTCCGTCTCGGCCGGC 3200
          |||   | ||| | |  |||| |  |  || |||| | ||  |||| || |||| ||
Db  1520 ATCGGCGGA-GACTGCTGATCGA-GACGCGAGATCGAGT-CCTCCATCATGCTGGGCGAC 1576

Qy  3201 GCCCAGGTGCAGGGGGCGTCCAGCATCGAGGCGTCCATCCTGGGGCGCCCGTCGTC 3260
          || || || ||| |  ||  || | ||| || | | |   ||||||||  || |||
Db  1577 TCCCGGCTC-ACGGCACGGCCCCCGGTCACCACTCGCTCCTGGGGCGGAACGTGACCGTC 1635

Qy  3261 GGCCCGGGCCCCCCGTCTCCCGGACTCGGTGATCGGCGACCAGCCAAGGTG 3320
          | |||||   ||||  | ||| ||  ||  |  | |||  | |||| ||| |
Db  1636 GCTCCAGCCCCGGTACCGGGACCGGTTCCGGCTCTCATCCTCGGTGACGACAGCAGGGTG 1695

Qy  3321 TATCTCACCCCATGACCACGACCATCCTCGTCACCCGGCGGGCTCATTCGCTCCG 3380
```

```
ACCESSION   AF080235
VERSION     AF080235.1  GI:4240402
KEYWORDS    .
SOURCE      Streptomyces cyanogenus.
  ORGANISM  Streptomyces cyanogenus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 34644)
  AUTHORS   Westrich,L., Domann,S., Faust,B., Bedford,D., Hopwood,D.A. and
            Bechthold,A.
  TITLE     Cloning and characterization of a gene cluster from Streptomyces
            cyanogenus S136 probably involved in landomycin biosynthesis
  JOURNAL   FEMS Microbiol. Lett. 170 (2), 381-387 (1999)
   MEDLINE   99126695
   PUBMED    9933932
REFERENCE   2  (bases 1 to 34644)
  AUTHORS   Westrich,L., Domann,S. and Bechthold,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) Pharmazeutische Biologie, Universitaet
            Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany
FEATURES             Location/Qualifiers
     source          1..34644
                     /organism="Streptomyces cyanogenus"
                     /strain="S136"
                     /db_xref="taxon:80860"
     gene            662..2137
                     /gene="lanE"
     CDS             662..2137
                     /gene="lanE"
                     /note="LanE"
                     /codon_start=1
                     /transl_table=11
                     /product="oxygenase homolog"
                     /protein_id="AAD13534.1"
                     /db_xref="GI:4240403"
                     /translation="MDAAVIIAGAGPAGLMLAGELRLAGVDIVLERLARTGESRGLG
                     FTARTMEVFDQRGLLARFGDVETSAAGHFGGIPLDFGLLEGAWKAAKTVPQSVTETSL
                     EEWAAELGADIRRGHELLSLREHGDAVEVEVRGPEGLRTLRAAYLVGCDGGRSTVRKA
                     AGFDFPGTAATLEMFLADVKGLDLLEPRMIGEEAAGGMVMVAKLPGGITRIIAGEYGTS
                     PRRATPPAYEEIAAVWKRLTGGDTHGEPVWLSAFGDAARQASEYRRGRVLLAGDSA
                     HIHLPAGGGQMNTSIQDSVNIQDSVNGLAGPDLDSYHSERHPVGERLLMNTRA
                     QGRLLLGGPEVQPLREVLDELVAFEVDRHLAGMVSGLEITYDVGAGTPLLGKRMPHL
                     ELVGEHAKSSSTELLHHGRCVLLDLADNARLREAPRLVGGVDVVTAEPHGVSEDSVLH
                     GTSAVLIRPDGHVAWAAPSSCHDLPMALIRWFGTPRVGRAF"
     gene            2220..2549
                     /gene="lanF"
     CDS             2220..2549
                     /gene="lanF"
                     /note="LanF"
                     /codon_start=1
                     /transl_table=11
                     /product="cyclase homolog"
                     /protein_id="AAD13535.1"
                     /db_xref="GI:4240404"
                     /translation="MHSTLIVAKMDPASSIDVAKLFGDFDRTEMPHRMGTRRRQLFSY
                     RGLYFHLQDFDENNGGELIEEAKTDPRFVAISQDLKPFIQAYDPETWRSPADAMATRF
                     YDWTASS"
     gene            2584..3858
                     /gene="lanA"
     CDS             2584..3858
                     /gene="lanA"
                     /note="LanA"
                     /codon_start=1
                     /transl_table=11
                     /product="B-ketoacyl-ACP synthase homolog"
                     /protein_id="AAD13536.1"
                     /db_xref="GI:4240405"
                     /translation="MGRRVVVTGIGVLAPGGIGAENFWSLLSEGRTATRGITFFDPSS
                     FRSQIAAEADFDAERSGLSPQEIRRMDRAAQFGVVTGREALADSGLDQAGLDPYRTGV
                     TIGSAVGATMGLDEEYRVVSDGGRLLDVDHHQYTPQHLYNHFVPSSFSAEVAWAVGAE
                     GPNTVVSTGCTSGIDSVGQAVELIREGSVDVMIAGATDAPISPITMACFDAIKATPR
                     NDEPEHASRPFDGTRNGFVLGEGSAVFVLELTAARARGANIYAEIAGYASRCNAFHM
                     TGLRPDGREMGEAIRVALDEARINPEAIDYINAHGSGTKQNDRHETAAFKLSLGEHAY"
```

| | | |
|---|---|---|
| Qy | 1696 | GAGATCTCCTCATGACCATGAAGATCCTGGTGACAGGCGGTGCCGGCTTCATCGGCTGCGC 1755 |
| | | ¦¦ ¦¦¦ ¦¦¦¦¦¦¦¦ ¦¦ ¦¦¦¦¦ ¦¦¦¦¦¦ ¦¦ ¦¦ ¦¦¦¦ ¦¦¦¦¦¦ ¦¦¦¦ |
| Db | 3381 | CCTACGTCCGCCGGCTCCTGTCGCCCCGGCCCCGGCGCGTCGCGGCTCGCCGGTGACCGTCCTCG 3440 |
| | | ¦¦¦¦¦¦¦¦¦¦ ¦¦ ¦¦ ¦¦¦¦¦ ¦¦¦¦ ¦¦¦ ¦¦¦¦ ¦¦ ¦ ¦¦¦ ¦¦¦¦ |
| Db | 1756 | ACTACGTCCGGCGCCTGCTCGCGGCGCGGCGGCTCGGTCGCGGGTCAC--TGCTGG 1813 |
| | | |
| Qy | 3441 | ACAAACTCACCTACGCCGGCAGCTCTCGCCCGCCGTGCGTGACCATCCCGGCC 3500 |
| | | ¦¦¦ ¦¦¦¦ ¦¦ ¦¦¦ ¦ ¦¦¦ ¦¦¦ ¦ ¦ ¦¦¦ ¦ ¦¦¦¦¦¦¦ |
| Db | 1814 | ACAGCCTGACCTATCGCAGGCGAATCCGGCCAACCTCGATCCGTCGCCGGCGACCAGCCGGT 1873 |
| | | |
| Qy | 3501 | TCAACCTTGCTCCAGGGCGACGTGTGCGACCGGCCTGGCCGGCGCGC 3560 |
| | | ¦¦ ¦¦¦¦¦¦¦¦¦¦¦ ¦ ¦¦¦ ¦¦ ¦¦¦¦ ¦ ¦¦¦¦¦¦¦ ¦ |
| Db | 1874 | TCGCCTTCGTCCACGTGACATCTGCGATCGACTCGTCGCCGCACTGGTCGCCCGAGC 1933 |
| | | |
| Qy | 3561 | ACGACGACATCGTGCACTTCGGCGACGTCGACCGCTCCAT---CACCGACA 3617 |
| | | ¦¦ ¦¦ ¦¦¦¦ ¦¦¦¦¦¦ ¦¦¦¦¦¦ ¦¦ ¦¦¦ ¦¦ ¦¦¦¦¦ |
| Db | 1934 | ACGACCAGATCGTCCACTTCGGCAGCCGAGTCACACGTCGACCCGTCGGTGCGCGTCGGGGGC 1993 |
| | | |
| Qy | 3618 | GCGGTGCCTTCACCCGCACCAACGTGCTGGCACCCAGTTCCTGCTCGACGCCGCGCTCC 3677 |
| | | ¦ ¦¦¦¦¦¦¦ ¦¦¦¦¦ ¦¦¦ ¦¦¦¦¦¦¦¦ ¦¦¦¦¦¦¦¦¦ |
| Db | 1994 | CGGGGAGTTCGTGCGCACCCAACGTGATGGGTACGCAAACGCTCGTGACACGCGGCGCCCTGC 2053 |
| | | |
| Qy | 3678 | GCCACGGTGTGCGCACCTTCGTCGCACGTGTCGCACGAGGGTACGGCGTCCCTCCCGC 3737 |
| | | ¦¦¦¦¦¦ ¦ ¦ ¦¦¦¦ ¦¦¦¦¦¦ ¦¦¦¦¦¦¦¦¦¦ ¦¦ |
| Db | 2054 | GCCACGCCCCCAGGATTTTCGTCCACATCTCCACGGACGGACGAGGTGTCGCACGTCCGGATG 2113 |
| | | |
| Qy | 3738 | ACGGGGCCGGCGGAGAGGCGACCCCCTGCTTCCGACCTCGCCGTACGGCGGTCGAAGG 3797 |
| | | ¦¦¦ ¦¦¦¦ ¦¦¦ ¦¦ ¦ ¦¦¦ ¦¦¦¦¦¦ ¦¦ ¦¦¦¦¦ ¦¦¦¦¦ |
| Db | 2114 | TCGGTGCCAGGCAGCACAGAGACCGATCCGTCGCACCCCAGTTCGGCGTACGCCGGCACGAAGG 2173 |
| | | |
| Qy | 3798 | CGGGCCTTCGGACCTCATGGCCGGCTCGCCCACCACGGCCACCGTCCGGTGA 3857 |
| | | ¦¦ ¦ ¦¦¦¦ ¦¦¦¦ ¦ ¦¦¦¦¦¦¦ ¦ ¦¦¦¦¦ ¦¦¦¦¦ ¦ |
| Db | 2174 | CGTCAAGCGATCGTTGGCGCTGTCGTACCACCGGCCACGCGCACGTCCGGATCA 2233 |
| | | |
| Qy | 3858 | CCCGCGTGTTCGAACAACTTCGGCCCCCCACCAGCATCCCGAGAAGCTCATACCGCCTTCC 3917 |
| | | ¦¦¦¦¦¦¦ ¦¦¦¦ ¦ ¦ ¦¦¦¦ ¦¦ ¦ ¦¦¦ ¦¦¦¦¦¦¦¦ |
| Db | 2234 | CCCGGTGCTCGAACAACTACGGCCACCATCAGCATCCGGAGAAGTCATTCCGCGCGTTTTG 2293 |
| | | |
| Qy | 3918 | TGACCAGCCTCCTGTCCGGCGGCACCGTTCCCCTCTACGGCGACGGGCACGTGCGCG 3977 |
| | | ¦¦¦¦¦ ¦ ¦¦¦ ¦¦¦ ¦¦¦ ¦ ¦¦¦¦¦¦ ¦¦ ¦¦¦¦¦¦¦ |
| Db | 2294 | TGACGACCCTGAGCGGTGCCGGGTGCCGCTCTACGGTGACCGGGAACCCTCGGA 2353 |
| | | |
| Qy | 3978 | ACTGGCTCACGTCGACGACACGTCAGGGCGTCGAACTCGTCCGCTGTCGGCGCCGGC 4037 |
| | | ¦¦¦¦¦¦¦¦¦ ¦¦¦¦¦¦¦ ¦¦ ¦¦ ¦¦ ¦¦¦¦ ¦¦¦¦ ¦ |
| Db | 2354 | ACTGGCTCACGTCGACGAGGACCACGTCGACGCGGTGGAACTGGTAAGGGAGGGAGGCTCCC 2413 |
| | | |
| Qy | 4038 | CGGGAGAGATCTACACACATCGGGGGCGCACCTGGGGGCCACCGTCCAACCTGGAGCTCACGCACC 4097 |
| | | ¦¦ ¦ ¦¦¦¦¦¦¦¦¦ ¦¦¦¦¦¦ ¦ ¦ ¦¦¦ ¦¦¦¦¦¦ |
| Db | 2414 | CGGGGCGAGATCTACACACATCGGCGGCGCACGGAGCTGTCCAACAAGGAGCTGACGGCTC 2473 |
| | | |
| Qy | 4098 | GGTTGCTCGACACTGTGCGGCGGCGGGCCGGACGCCGATCGTCCACGTCGAGAACCGCAAGG 4157 |
| | | ¦¦¦¦¦ ¦¦¦ ¦¦¦ ¦¦ ¦¦¦¦ ¦¦ ¦¦¦¦ ¦¦ ¦¦¦¦¦ |
| Db | 2474 | TCCTGCTGACGGGTGGGCCCGGTGGGACTCCGTGGGACTCGTCGCCCGACCCCGGCCAAGG 2533 |
| | | |
| Qy | 4158 | GGCACGACCGGCGCTACGCGGCTACCGTGCGTCGACCACAGCAAGATCACCGCGGAACTCGGTTACCGGC 4217 |
| | | ¦ ¦¦¦¦¦¦¦¦¦¦ ¦¦¦¦¦¦¦ ¦¦¦¦ ¦ ¦¦¦¦¦ ¦¦¦¦¦¦ |
| Db | 2534 | GACACGACCGGCGCTACGCGGCTACCGTGCGACCACACGGGAAGAGATCACCGCCGGGTACCGGC 2593 |
| | | |
| Qy | 4218 | CGCGCACCGACTTCGGCGACACGCCTGCCGCAAGCCGAAGTGGTACGAGCGGCACGAGG 4277 |
| | | ¦¦¦¦¦¦¦¦ ¦¦¦¦¦¦ ¦¦ ¦¦¦¦ ¦¦ ¦ ¦¦¦¦¦¦ |
| Db | 2594 | CACGACCGGGACTTCCCGGGACTCGCTCGCCGAGACAGTGGCCTGGTACCGGCGACCGCGCT 2653 |
| | | |
| Qy | 4278 | ACTGGTGGCCGTCCCCTGCTCG 4298 |
| | | ¦¦¦¦¦ ¦¦ ¦¦ |
| Db | 2654 | CCTGGTGGCACCGCTCACCG 2674 |

```
RESULT 10
AF080235
LOCUS       AF080235            34644 bp    DNA     linear   BCT 08-FEB-1999
DEFINITION  Streptomyces cyanogenus landomycin biosynthetic gene cluster,
            complete sequence.
```

```
        RTPISSIKSMVGHSLGAIGSIEIAASLLAMENHVVPPTANLHTPDPECDLDYVPLVAR
        EHTTDHGPDGRQRIPVGFQSAMVLARPERSAA"
gene    3855. .5075
        /gene="lanB"
CDS     3855. .5075
        /gene="lanB"
        /note="LanB"
        /codon_start=1
        /transl_table=11
        /product="chain length factor homolog"
        /protein_id="AAD13537.1"
        /db_xref="GI:4240406"
        /translation="MTARVVITGNRIAAPNGFGVEDYWAATRVGKSAIGRITRFDPTQ
        YPARLAGEIRGFDARDHLPGRLLIPQTDRMTQLALVATDSAFEDAGVKPGDIPEYDMGV
        VTASTAGGFEFGQONELAQALWSKGSQHVSAYQSFAWFYAVNSGQISIRNGMRGHSAVVV
        SDQAGGLDAIAQARRQIRKGSKLICSGGVDASICPWGWVAQLANGRVSTSENTERAYL
        PFDADASGVYPGEGGALLILEDAERARQRGAENVYGEIAGYGSTFDPRPGSGREPRLR
        RAIEIALADADAEPSDIGVVFADAAGTPELDRVEAEALIEVFGAAGVPVTAPKTMTGR
        LYSGAAPVDVVTAVLAMREGLIPPTVNVSLSPEYDIDLVTAQPRTARVRNALVVARGY
        GGFNSAMVVRGTDR"
gene    5135. .5404
        /gene="lanC"
CDS     5135. .5404
        /gene="lanC"
        /note="LanC"
        /codon_start=1
        /transl_table=11
        /product="acyl carrier protein homolog"
        /protein_id="AAD13538.1"
        /db_xref="GI:4240407"
        /translation="MSQPEFTVEDLKRILVEGAGADEGVDLDSDILDTDFESLGYESL
        ALLETAGRIEREFGITLDDEVFIDNPTPRELVAAVNAKLPGSLAA"
gene    5443. .6228
        /gene="lanD"
CDS     5443. .6228
        /gene="lanD"
        /note="LanD"
        /codon_start=1
        /transl_table=11
        /product="reductase homolog"
        /protein_id="AAD13539.1"
        /db_xref="GI:4240408"
        /translation="MSQQDRVALVTGATSGIGQPPGCSPPRGTGCSRRGNAENVAET
        VKQLQLQDEGLDVDGTALDVRSADGVRAFVQSAVDRFGTVDVLVNNAGRSGGGVTADI
        ADELWDDVIATNLTSVFRMTREVLNTIGGMRGKSRGRIINIASTAGKQGVVLGAPYSAS
        KHGVVGFTKALGNELAPTGITVNAVCPGYVETPMAQRVRQGYAAAYNATEDAILEKFQ
        AKIPLGRYSTPEEVAGLVGYLASDTAASITSQALNVCGGLGNF"
gene    6276. .7235
        /gene="lanL"
CDS     6276. .7235
        /gene="lanL"
        /note="LanL"
        /codon_start=1
        /transl_table=11
        /product="cyclase homolog"
        /protein_id="AAD13540.1"
        /db_xref="GI:4240409"
        /translation="MTVREVEHCDHRARPPAAEVYRLIAEVENWPRVFPPTIYVDQVE
        RGTPGTGVSEERIRWATANGTAKNWTSRRTLDPQALRITFROEVSAPPVAAMGGTW
        IIEPLAADESRIRLLHDYRAIDDDPDSLKWIDEAVDRNSRSELAALKTNVELAHATEE
        VTFSEDTVRINGSAKDVYDFIDEADKWEERLPHVASVRLEEISPGLIQILEMDTRTKD
        GHTHTKFYRVCFSHQKIAYKQVTLPALMTLHTGYWTFTENEDGVTASSQHTVVLNTE
        NITRYLGPDAGIAEAREHVQTALSANSRATIGLAKDYAEQQQR"
gene    7238. .8806
        /gene="lanM"
CDS     7238. .8806
        /gene="lanM"
        /note="LanM"
        /codon_start=1
        /transl_table=11
        /product="oxygenase homolog"
        /protein_id="AAD13541.1"
        /db_xref="GI:4240410"
```

```
        /translation="MAAKHTDTDTVIVVGAGPVGLMLAGELRLGGADVVVVERLSTPTT
        ESRASTLHARTMEILDSRGLLIDELVPVPNDVMGHFGGIPLDLTIPSAYPGQWKVPQTR
        VEELLQGWATGLGAVVLRGHELRGLTVGQGHVAVDVLGPDGPVRMRGRYVVGCGEQS
        AVRRLSGIAFPGEDASRELIRADVAGIDIPPRRFQRLESGLAIAARRPDGVTRVMVHE
        FGRTAEPRSAEPFSEVAEIWKRVTGEDISGGEPLWVNAFGNASRLAERYRDGRVLLA
        GDAAHQQMPVGGQALNLGLQDAVNLGWKLAAQVTGRQPDGLLDSYHTERHAVGRRTLS
        NIRAQSLLLLGGGEVEGVRGVFAEITAMEDVRTRLAGMISGLESRYEVGPGHHPLLGA
        RVPHVELAGEQGVTVTTALLRPGRAVLIGLSGDPLRGARLRAAATPWFPRVTVTLAEP
        VDHSAPLADVDAVLARPDGHVVWTGSAGEAELHAVLRRWFGEPEPAGTAGTRFRAHGS
        STPLGADSPAGPPW"
gene    8803. .9471
        /gene="lanN"
CDS     8803. .9471
        /gene="lanN"
        /note="LanN"
        /codon_start=1
        /transl_table=11
        /product="reductase homolog"
        /protein_id="AAD13542.1"
        /db_xref="GI:4240411"
        /translation="MTHAGGGAGRAAAMRLAAAGALVAIHHTGEQAADDVVRAIGEG
        GGCAFALGAELGGYGNVORLFADLEAELTDRTGRADLDILVNSADDVDGVAPEHATPE
        EFDRLVAAGAKAPVFVIQRALRSLADGGRVINISRGAVEPVEPKAVAVAMAKGAVETL
        ARHYAVALAPRGITVNSVAPGTTDDGERVGTVVAFLAGDDAGGITGAVFDADGSPVHG
        RPAG"
gene    9626. .10207
        /gene="lanO"
CDS     9626. .10207
        /gene="lanO"
        /note="LanO"
        /codon_start=1
        /transl_table=11
        /product="reductase homolog"
        /protein_id="AAD13543.1"
        /db_xref="GI:4240412"
        /translation="MSQEPLRLALLIGSNREARFGEVVGDWFAGVAKRRGDLTVDID
        LADHELP"VLSQNPSPRVAAELEKVSPRLAEADAYVVVTPEYNHSYPAALKNLIDWHH
        TQWQAKPVGFVSYGGLSGGLRSVEHLRPVFAELHAVTVRDVVSLHMAWERFGPDGAPK
        DEGPAQAAKVLLDQLVWMGNALREARAKTPYGI"
gene    10210. .11784
        /gene="lanP1"
CDS     10210. .11784
        /gene="lanP1"
        /note="LanP"
        /codon_start=1
        /transl_table=11
        /product="decarboxylase homolog"
        /protein_id="AAD13544.1"
        /db_xref="GI:4240413"
        /translation="MTSVEEGVPDIQDRLDELRELRKQAREGLDPSATERQHAKGKLT
        AYERIELLLDEDSFNEVEPLRRHRAQGFGLEAKKPYTDGVITGWGTVEGRTVFVYAHD
```

```
Query Match           12.7%;   Score 867.2;   DB 1;   Length 34644;
Best Local Similarity 65.3%;   Pred. No. 7.2e-58;
Matches 1325;   Conservative  0;   Mismatches 688;   Indels 15;   Gaps  3;

Qy   2264  CGAGGTGAAGGCCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCA   2323
           ||  ||||||| |||||||| ||  | |||||| |  | ||  |  ||||| ||
Db  12082  CGACATGAAGGCTCGGTCCTGTCGGGCGGTCGGGCGGAACCCGGTTGCGCCCCTTCAGCCA  12141

Qy   2324  CACCGCCGCCAACGCAGCTGCTCCCCATCGCCAACAAGCCGGCGTCGTCCTTCTACGGCTGGA   2383
           |  |   ||||   | | | ||   |||  |  ||| |||| || ||  ||||||||
Db  12142  CTCGATGCCGAAACAGCTCATACCGATCGCCGATCGCCAACAGCCCGTCCTGGAGCACGTCCTCCG  12201

Qy   2384  GTCCCTCGCCGCCGGGGTGTCCGGGAGGCCGGCGGCTCGTCGTGGCCGGCCCGGGGA   2443
           | || ||  | ||   | | || ||| |  ||||  ||||| ||| |    |||
Db  12202  GAACGTCAGGGACCTCGGCGTGACGGAGATCGGCGGTCATCGTCGGCGACCGGCCGAGCA  12261

Qy   2444  GATCCGCGAACTCACCGGCGACGGCACCGCCGGCGTTCGGGTTACGCATCACCTACCTCCACCA   2503
           |||||     ||| ||||||| ||||||||  || |  ||    ||| |||   || ||||| ||
Db  12262  GATCCAGGGGGTCATCGGCGAGCGGCGCTCCGGCGTCTGGAGATCCACCTCACATCCCGCA  12321

Qy   2504  GCCCCGCCCGCGCTCGGTCTGCGCACGCGGTGCGCATCGCCGCCGGCGGCTTCCTGGGCGACGA   2563
```

```
                        |||||| ||||| ||| |||  |||||||| |||||| |||||||
Db 12322 GGAGGCGGCCGGCTTCGGCCTGAAGCTGCCCACTGCGCCCGGCCCCTTCCTCGGCCGACGA 12381                Db 13393 TTCGCCGCGAGTCGCACGTGGACCGCGTCTCTGGAGTCCGGAGGAGTTCGTCCCACACC 13452
Qy  2564 CGACTTCCTGCGTACCTCGGGGACAACTACCTGCCCCAGGGCGTCACCGACTTCGCCCG  2623                Qy  3638 AACGTGCTGGGCACCCAGTCCTGCGCCACCCAGGCCGCGCTCGGACGCGTGCCGCACCTTC 3697
              ||||| ||| |  |  ||||||  |||||||||  ||                                                     ||||| |  |||||||  ||||  ||   |||||
Db 12382 CGACTTCGTGATGTACCTCGGCGACAACATGCTGCCCGGCCGGGCGGGGATCGCCGA 12441                Db 13453 AACGTCACCGGCACTGCTGCACGGCGGTGCTCGCCGCAGTCAAGGCGTCAAGCGCGTC 13512
Qy  2624 CCAATCGGCGCCGGATCCCGGCCGCCGCCCGGCCGGGTCGCGCGGGATCGGCCCGGTC  2683                Qy  3698 GTGCACGTCTCCACGAGAGTGTACGGCGGTCCCTCCCCGCACGGGGCCGCCGGGAGAGC 3757
            | || |||| |  ||||||| || |||| |||  |  |||||||||||||||||                                     || || ||||||| || ||||||| ||||||  |||
Db 12442 CGAGTTCACGGCGCCGCCCGGCCCCAGGTCGTGTCCAAGAGGCGCCACCGACCCGTC 12501                Db 13513 GTCCAGTGTCCAGACGAGGTCACGGCTCCATCGACGAGGGGGATCCTGGACCGTCGAGGAG 13572
Qy  2684 CGAGTTCACGGCGCCGCCCGAGGAGTCGACGGAGGAACGTGCTGCGGTTGGAGGAGAAACC  2743                Qy  3758 GACCCCCTGCTTCCGACCTCGCCGTCACCGGCGTCGAAGGCGGCCTCGGACCTCATCGCCG 3817
              ||||| |||||||||   |||  |||  |||  || |||| |||||                                            |||||  ||||| || |||| ||||| |||||||
Db 12502 TGCCTTCGGGGTCGCCGAGCTGCCGCCGGACGGCGCAGGTCGAGCGCCGTGGTGAGAAGCC 12561                Db 13573 TGGCCGCTGGCGCCCCAACTCCCGTACTCCGGCTTCCGGCTTCTCCGTCCGCCTGCTCGCA 13632
Qy  2744 CGACGTCCCGCCAGCTCGTCATCGGCGTGTACGCCCTTCAGCCGCCGTCCA  2803                Qy  3818 CTCGCCCCACCACCGCCACCCAGGCCGCTCGCCGGCGGTGACCGCCGTGTTCGAACAACTTC 3877
            |  |  |||  |||| |  |||||| ||||||||||||||||||                                             |  | |||||  |  ||| ||| ||||| ||||||||||| ||||
Db 12562 GCAGGAGCCCAAGAGCCGGATCTCGCGCGATCTCGGCCGTCTACTTCTTCACCCCGGCATCCA 12621                Db 13633 CGCTCGTACTGGGCGCACCCGGGCCTCGACCTCGTCCATCACCCCGGTGCTCCAACAACTAC 13692
Qy  2804 CGAGGCGGTACGGCGCCATCACCCCGTCCGCGGGCGAGCGTGGAGATCACCCACGCCGT  2863                Qy  3878 GGCCCCCACCAGCATCCCGAGAGCTCATACCGCGCCGTTCCTGACCAGCCTCCTGTCCGGC 3937
             ||| |||||| |||| |||||||  ||||||| ||||| || |||||||                                          ||||| |||||| || |||||||  || ||||||||||||||
Db 12622 CAAGGCCGGTTGCCGCCATCGAGCCGACCTGCGCGGCGAGATCACCGACGCCCAT 12681                Db 13693 GGCCCGTACCGACCCGGAGAGCTCATCCCGCTGTTTGCTCACCAACCTGCTGGAGGGT 13752
Qy  2864 GCAGTGTGGATGATCGACCGGCGGCTGCGCGTACGGCGCCGAGACCACCACCCGGCCCTGCGCG  2923                Qy  3938 GGCACCGTTCCCCTTCTACGGCGACGGGCCGCACGTGCCGCACTGGCTGCACGTCGACGAC 3997
              |||| |  |||    ||  ||| |||   |||||||||||  |||||||||                                        |||||||||||||||| |||||||||||||||  |||||||||||
Db 12682 CCAGTGGCTCGTCGCCAACGGCAGCAGGTTCAAGGCCAACGAGTACGAACAGGGCTACTGGA 12741                Db 13753 GAGCAGGTGCCGCTCTACGGCGGCGACGGGCCGCAACATCCCGCAGTGGCTGCACGTCGACGAC 13812
Qy  2924 CGACACCGGACGCGCGGGAGACCTGGAGGTCAACCGTCACGTCCTCTGACGGGACTGGA  2983                Qy  3998 CACGTCAGGGCCGTCGAACTCGTCGCCGACGTGTCGGGGCCGGCGGGAGAGATCTACAACATC 4057
             ||||||| | ||  ||||||||  ||||  ||||||| |||||||||||  |||||||                                 ||| || |   |||||| |||| ||||||||  ||| |||||||||||
Db 12742 GGACACCGGCATCGAGGGAGACGTCCTGGAGTGCAACCGGCACGTCCTGCGGCGAGCTGA 12801                Db 13813 CACTCCCGTGCGATCGACTTCGTCCGTCGACGTCCTGAGGGTCCGGGGAGATCTACAACATC 13872
Qy  2984 GGGCCGCATCGAGGGGAAGGTCGACGCGCAGCAGCATGCTTGGTCGGCCGGGTGGC  3043                Qy  4058 GGGGCGCGGCACCTGCTGCCCAACCTGGAGCTCACGCTGCTCGGCCTGCTGCTGCGCGGC 4117
            ||| |  ||||||  |||||||  ||||||||| ||||  || |||||||||                                       ||||||| | ||||||  ||| |||||||||||| |||| || ||||||||||
Db 12802 GCCGGCCACCGGCGAGGTCGCAGCGCGCAAGCACGCGTCGCATCCGGCACGCGTCGTGGTGA 12861                Db 13873 GGCGGCGCACGCACGAGCAGCCCATCACGAGCGGTCATCACGCGCTCGCTCGACGGGA 13932
Qy  3044 CGAAGGCGGCCATCGAGTCGTGCGGGGTCACACGTGGTGCGCCCGGTGTGATCGGGCGGGTGC  3103                Qy  4118 GCGGGGCGCCGGGCCCGGAGAGCATCGTCCACGTGCCACGTGCTCCGAGAACCGCAAGGGGA 4177
              |||||| |  |||||||| | |||| |||||||||| | ||||||| | |||||||||                                |||  |||  || ||||| || |||||||||||| |||||||||||||||
Db 12862 AGCGGGTCCAAAGTCATCGGCTCCGGCTATCGAGGGCGGCCCCGTACACCTCCATCGGGGA 12921                Db 13933 CAGGACTGGTTCCAAGGTCCGTCACGTCCCGACCGGGAGATCGCCCAAGGCGCACGACCTGCGCTACTCC 13992
Qy  3104 CGTCGTCAGCGAACTCCAGTGTCGGCGGCCCCGTACACCTCCATCGGGGAGGACTGGGTGA  3163                Qy  4178 GTCGACCACGACAAGATCACCGGGGAGGAACTCGGTTACGGCGGTTCCGGCACTTCCGCGACC 4237
             |||  ||| | ||  |  |||||||||||||||||| |||||||||||||                                         ||||  ||| || || || || |||||| |  |||||||||||||||||||
Db 12922 GGTGGTGAAGGACAGCGCACATCGGCCCGGACGACGCACTGCACGCACTCCCATCGGCGACAGCGCTGAC 12981                Db 13993 CTCGACCAGTCGAAGATCGCGAAGATCCGGCGGAGGAGCGGCATCCCCGCCATCACCTTCGAGCAG 14052
Qy  3164 GGACAGCGCCATCGAGTACTCCCGGCGGCGGCCCAGTTCGAGGGGGCGTCCCG  3223                Qy  4238 GCGCTGGCCGACACCGCGAGTGGTACGAGCGGCACGAGGACTGGTGG 4285
              |||||||||||||| |  | ||||||||  || ||||||||||| | |||                                       ||||| ||| ||| ||||| |||||||||| ||||||||||||||
Db 12982 CGGCACGACGAGCATCGAGTACTCCGATCGCCGGCGGCCGCGTCGTCGCCGCCCCGTCTCCCGCGA 13041                Db 14053 GGCCTCGCCGACCACCTCGCCGTGGTACCACGACGACAACCCCGGCTCGTGGTGG 14100
Qy  3224 CATCGAGGCGTCCATCGCCGCCGCCGCCGTCGTTCGGCGGCCCCGTCTCCCGCGA  3283
            |   |||||||||||| || |||||| |||  ||||||||||||||
Db 13042 CCTGCACAGTCCGTCATCGGCCCGCTCGACCGGTCGTTCGGTGCGGGAGCAGCCGGAGCACCCG 13101
Qy  3284 GGCTACCGACTGGTGATCGGCGACCACCAAGGTGTATCTCACCCCCATGACCACGACC  3343
            ||| ||||| ||| ||||||| ||||||||| |  |||| || |||||| |||||||||
Db 13102 CCATCACCCGGCTTCGTCGTGCTGGTGACCACCACCCGTCGAGTCGGACGGCCATG---AAGAAG 13158
Qy  3344 ATCCTCGTCACCCGGGAGAGCCGAGCGGGCCTTCATTCGCCTCCGCCCTACGTCCTCGTGTCG  3403
            |||||||| ||||| ||||||||  || ||||||| |||||||||| ||  | ||||||||
Db 13159 ATCCTGATCACCCGGTGCCGGCGGCTTCATCGGTTCGCACTACGTCGCACCCTCCGGTGAAC 13218
Qy  3404 CCCGGGGCGCCCCGGC------GGGCTCGCGGTGACCGTCGCCCAGTTGCGAGGGGCGTCCCG  3457
            |           |||||            ||   |  |||||||||||| ||||||||| |||
Db 13218 GACGGGTACGAGGAGGACTCGGAAGGGCGGCGCCGCTCCCGCCTCGGACAAGAGCTCACCTACGCC 13278
Qy  3458 GGCAGCCGACTGCGCTCGCCGGGCTACCTCCACCTTCGTCGTCAGGGC  3517
            ||||||||||| |     | ||||| ||||  |  ||||||||| ||||
Db 13279 GGCAACCGGCCGCAGACCTGCGCCGCGCCC-----CACCCGCGGCTAGCCGGTGTCTCCAGGGGC 13332
Qy  3518 GACTGTGTCGACACCCGGCGACCGCCTGCTCGCTCGCAACTGCTCCGGGGACATCGTGCAC  3577
            |  |||||||  |||| |||||||||||| |||| ||| |   | ||||| |||||||| |
Db 13333 GACATCTGCGACTTCGAGCTGCTGCTCGAACTCGTCGCTGCCGGGACAGCGGTCCGCTCCAC 13392
Qy  3578 TTCGCCGCGAGTCGCACGTGGACCGCGTCTCTGGAGTCCGGAGGAGTTCGTCCCACACC  3637
            |||||||||| || ||||||||||||||||||||||| ||||
```

```
AUTHORS    van Wezel,G.P., White,J., Bibb,M.J. and Postma,P.W.
TITLE      The malEFG gene cluster of Streptomyces coelicolor A3(2):
           characterization, disruption and transcriptional analysis
JOURNAL    Mol. Gen. Genet. 254 (5), 604-608 (1997)
MEDLINE    97340948
PUBMED     9197422
FEATURES             Location/Qualifiers
     source          1..5010
                     /organism="Streptomyces coelicolor"
                     /db_xref="taxon:1902"
     gene            complement(194..1265)
                     /gene="malR"
     CDS             complement(194..1228)
                     /gene="malR"
                     /codon_start=1
                     /transl_table=11
                     /product="repressor of malE transcription"
                     /protein_id="CAA68975.1"
                     /db_xref="GI:1524332"
                     /db_xref="SWISS-PROT:P72396"
                     /translation="MTTRLADIAAQAGVSEATVSRVLNGKPGVAATTRQSVLAALDVL
                     GYERPVRLRQRSEGLVGLIITPELENPIFPALAQVIGQALTRQGYTPVLATQTPGGSTE
                     DELTEMLVDRGVAGIIYVSGLHADTTADMQRYERLRAQGVPFVLVDGFSSQVQAPFIS
                     PDDRAAMSLAVTHLVSLGHTRIGLALGPKRFVPVQRKIEGFVRTVQDLGLSAETVEK
                     ELVQHSLYTLEGGQAAASALIGRDCTAVVCASDMALGAIRAARQLGLDVPKDVSVVG
                     FDDSPLIAFTDPPLTTVRKPVPAMGQAAVRTLLEEIGGTPAPHSEFVFMPELVVRGST
                     ASAPGERGRP"
     -10_signal      complement(1237..1242)
                     /gene="malR"
     -35_signal      complement(1260..1265)
                     /gene="malR"
     gene            1491..2891
                     /gene="malE"
     -35_signal      1491..1496
                     /gene="malE"
     -10_signal      1513..1518
                     /gene="malE"
                     /note="putative"
     mRNA            1525..2891
                     /gene="malE"
     RBS             1610..1614
                     /gene="malE"
     CDS             1620..2891
                     /gene="malE"
                     /codon_start=1
                     /transl_table=11
                     /product="putative maltose-binding pootein"
                     /protein_id="CAA68972.1"
                     /db_xref="GI:1524333"
                     /db_xref="SPTREMBL:P72397"
                     /translation="MRGIAATALVASLALAATACGGDGDSDSESGGPVTITYWDTSN
                     ATNEAPTYKALVKEFEAANKGVKVNFVNVPFDQAQNKFDTRAGSKGAPDVLRSEVGWT
                     PAFAKKGFFLPLDGTEALAEQDKFQPNLIEQAKYEGKTYGVPLVTDFLAFVNKELFE
                     KAGVEAPKTWDDLKKRAATVKDKTGVDGYWASTAGYYAQPELYGEGTDTVDADAKKVT
                     VNAAAAKKGYGTWLSLFDGKGLHKADTTADAYAHIQEAFVSGKVASIIQGPWEITNEY
                     KGTAFKDKNNLGIATVPAGSTGKAGAPTGGHNLSVYAGSDKAHQEAALKFVNFMTSAK
                     SQETVALKNSTLPTRDDAYTAKVKADPGIAGFQTVLPAAQPRPALPEYSSLWTPLDDE
                     LPQIAGGKKSLDEGLGDAETAIAKLVPDFSK"
     gene            2997..4014
                     /gene="malF"
     RBS             2997..3000
                     /gene="malF"
     CDS             3010..4014
                     /gene="malF"
                     /codon_start=1
                     /transl_table=11
                     /product="putative maltose permease"
                     /protein_id="CAA68973.1"
                     /db_xref="GI:1524334"
                     /db_xref="SPTREMBL:P72398"
                     /translation="MTVAIDRATGKKRGDRAPRPGLGRRLKDGYQKHWYAYAMIAPVA
                     VVIGDLVLYPLGYGLYLTLTDANSLNSARTIGVNEIEATYKFVGLDNYADILWGPTAY
```

```
     gene            4018..4936
                     /gene="malG"
     RBS             4018..4022
                     /gene="malG"
     CDS             4028..4936
                     /gene="malG"
                     /codon_start=1
                     /transl_table=11
                     /product="putative maltose permease"
                     /protein_id="CAA68974.1"
                     /db_xref="GI:1524335"
                     /db_xref="SPTREMBL:P72399"
                     /translation="MSTSTAPTKAPAEQIAAPTAPPRKVRGRGERSRAASLASHGVLT
                     LASLVAVFPVAWLVYLSLGPDKNDYLHPGRIWSKMTFDNYAFVLQDTNFFDWLKSSLI
                     VSLGTTVICVLVAATTATRSRMRFPGYRKLMWVLLVTQMFPIAVLIVPMYQILSKIQL
                     IDNYFGLILYCSTAVPYCAWLLKGYFDTIPFEIDEAGRVDGLTPFGTFFRLILPLAR
                     PGLAVAGYSFIATAFGEVAFASTMLSDFKYTFAVGLQSFVSEHDAQRNLMAATAVLV
                     AIPVSAFFLVQRNLVTGLTAGGTKG"

BASE COUNT      807 a    1879 c    1572 g     752 t
ORIGIN

Query Match              12.5%;    Score 855;    DB 1;    Length 5010;
Best Local Similarity    63.0%;    Pred. No. 1.2e-56;
Matches 1683;    Conservative   0;    Mismatches 800;    Indels 188;    Gaps 16;

Qy    4335  CCCCGGCCGGCACACCGCCGCCCCCCGCCGGTGGCCGGCCGGTCCGTGAGCCGG  4394
            || ||||| ||  |||  |  ||||| | |||||  || ||| |||| ||||
Db     142  CCTCAGGTGGAACCGACCCGCATCCGCCCCCGGAGCCCCCGACGGGAATCAG--GG   199

Qy    4395  GGGGCCGGCCGCCCCCCGGCGGCCGGGCGGCCGGTCCGGCCCGCGTTCCGGGCATGAA  4454
            ||| ||  ||  |||||| | | ||||  || || ||| |||| | ||||| |||||
Db     200  GCGGCCGGCCTCGCCCGGCGCCGGCGTCGAACCGGCCACCACCAACTCCGGCATGAA   259

Qy    4455  GACGAATTCGGTGCGCGGCGGCGGGCGTTCCGCTCATCTCCTCCAGCAGTGCGTCCACGGC  4514
            |||||| ||| |  ||||| | | || |||| ||  ||| |||  |||| ||  ||  ||
Db     260  CACGAACTCGCTGTGCGGCGGCGCGGGCGGGCCCCGATCTCCTCCAGGAGCGGCACCGGC   319

Qy    4515  GACCTGCCCATCGCCTTGACGGGCTGTCTGATGGTGGTCGAGACGATGTCACCGGCAACGC  4574
            |||||||||| | ||  ||||| | || |||  |||| |||| | ||  |||||| |||
Db     320  CGCCTGCCCCCATCGCCGGGGACCGGCTTGCGACGTCGGTGAGCGGCGGGTCGGTGAAGGC   379

Qy    4575  CATGAGCGGCGAGTCGTGAAGCCGACCACCGAGATGTCGCTGCGTGGCACATGACGGCGGT  4634
            ||  ||||| ||||||| || ||||| ||| | ||| |||||| | ||||||||| |||
Db     380  GATCAGCGGCGAGTCGTCGAAGCCCAGCGACCAGCTGGCAGTGTCCTTGGGCACGTCCAGGCCCAG   439

Qy    4635  CCGGGCCGGCGCGCCGGGCCATCATGTCCGCAGGGGAACAGCGA  4694
            | || || || |||||| | ||||  ||  |||| |||| || |
Db     440  CTGCGGGGCCGGCGCGTATCGCCGGCGCGGGGGCCTGCCCCGT   499

Qy    4695  GCAGCCCAGGTCGATCAGCGCGACGCGGGAGGCCGCCGCCGGAGAACAGCGA  4754
            | ||||| ||||| ||||||| | ||||| ||||| || ||||| | |||
Db     500  GCAGTCCCGGCCGATGAGCGGCGAGGCCGCGCCTGGCCCGTACGCGA   559

Qy    4755  GTGCTGCACGAGCTCCTCGGACTCCGC---GCCGACACTCCCAGTGCGTCCCACGCGCGCC  4811
            ||||| |  |||||| ||||| ||  |   |||| | ||| |||| | || | |  ||||
Db     560  GTGCTGGACCAGCTCCTTCTCCACCGTCTCGGCGCCGCAGAAGCGGGCGGCCCCGAGCGGT   619

Qy    4812  GGCCCGGAACCCGTCGATCTTCCGGGCTGCACCGTGCACCGGCACAAGCGGGCGGCCGCGAG  4871
            || | | || |||||||||| |||| || |||  | || |||||| ||| || |||||
Db     620  GCGTACGAAGCCCTGAGTCGATGCGCTGGCCACCTGGCACCGGCACGAACCGCTTCGGCCCCGAGGGCCAG   679

Qy    4872  GCCGACGGCGCTCGTGCCCCAGGTCGCCGCCAGTGCGCCCAGGCGCGCCATCGCGGCCCCG  4931
            | |||   ||| ||| | ||  |||| || | ||| |||| || |||||||||||||
Db     680  TCCGATCCTGTGTGCGCGAGGACGAGGTGGGTGACCCGGGCTCATCGCCGCCCCG   739

Qy    4932  GTGTCCGGGGGAGACGAAGGTGCCTCGATCCGGGAGGCCGAGACGGTCGACGAGACGAA  4991
            ||| ||||||||||| ||| | ||  |  ||| |  || |||| | ||||||||||||
Db     740  GTCGTCGGGGGAGATGAACGGCCGCCACCTGGCACCGGCACGAACCGCTTCGGCCCCGAGGGCCAGACGGACGAA   799
```

```
Qy   5932 TCAAGAACGCCGGCGGGCGGCAACTCCGGTGCCCCCGACGTGATGCTGACCGAGGTCGCCT 5991
          |||  |||||| |||  ||| ||||||||||||||| ||  || ||||||||||||||||
Db   1864 TCGACACCGCGCCAGGCTCCAA---GGGCGCCCGGACGGTGCTGCGCTCGAGGTGCGGCTG 1920

Qy   5992 GGGTGCGCGGACTTCGCCAGCATCGGCTACCTGCCCCTCGCCGTCGACGGCACCGCCCTCG 6051
          |||||||||||  || |||||| |||||  ||| ||||| |||||||||| ||||||||||
Db   1921 GGACGCCCGGCTTCGTCGCCAAGAAGGGCTTCTTCCTGCGCTCGCTCGACGGCACCGCCCTCG 1980

Qy   6052 ACG--ACGGGTCGGACCACCTTCCCAGGCGGCGGCACCCGGTACGAGGGGAAGACCT 6108
          |||   ||||| || || |||||||| || |  ||||| || |||||||||||||||
Db   1981 CCGAGCAGGACAAGTTCAGCCAACCTAGTCGAGCAGGCCAAGTACGAGGGCAAGACCT 2040

Qy   6109 ACGCGGTCCCCAGTGATCGCACCCTGGCCTCTTCACAACAAGGAACTGCTGACGA 6168
          |||   ||||| |  |||||||||||||||  ||  |||||||||||||||||
Db   2041 ACGGCGTGCCGCTGGTCACCGACGTCCCGGGCCTCCCGAGCTGAACGACGCCGCCGAGATCA 2100

Qy   6169 AGGCCGGTGTCGAGGTGCCGGGCTCCCTCGCCGAGTCGGCGCCCGAGACGGCGCCGAGATCA 6228
          ||||||| ||||||||||||||||||||  || ||||||||||
Db   2101 AGGCCCGGCGTCGAGGGCGTGACCGGTCCCCAAGACCTGGGACGACCTGGGACGACGCCGCCGCCACCGTCA 2160

Qy   6229 CCGAGAAGACCGGCGGCGAGGCGGCCGACCTGGCTCGACGAGGAACAAGACCGTCACGGTCG 6288
          ||||| |||||||||||  |     |||| || ||||||||||||| ||
Db   2161 AGGAACAAGACCGGCG---TCACGGCTACTGGGCCTCCCCACGCGCCGCTACTACGGCAGC 2217

Qy   6289 CCTACCTTCTACGGGAGGGGGCGACCTGGTCGAGGAGAACAAGACCGTCACGGTCG 6348
          |||  ||| ||| ||||| ||||| || |||||||||||||||||||| ||||||
Db   2218 CGTTCCTGTACGGCGAGGGCACCGACGACGGTGACGCGCCGCCAAGAGGTCACCGTCA 2277

Qy   6349 ACGACGAAGCCGGTCGTGCGCGGCCGACTCGTCGACAGCAGCAAGCAAGGCGG 6408
          |||| |||||||||| ||||||| ||||| || |||||||||||||||||||||
Db   2278 ACGCCGCCGGCGGCCCAAGAAGGGCTACGGCCGCCACCTGGCTGAGCCTCTTCGACGACGACGCAAGGGGCC 2337

Qy   6409 CCATCACCGACG------CGTCCGACGGCTGAACAACATGCGAACAACTCCGAAGTCGG 6462
          || |||  ||||    ||||||| |||||||||| || ||||||||| |||||||||
Db   2338 TGCACAAGGCCGAACTGCGAGACGCCTACGCCGACGCCTACGAGGCGGTTCGTCAGCG 2397

Qy   6463 GCAAGGTCGGCCATGATGGTCAACGGCCCATGCAGGACGGTCAAGGCGACGGGAGCCC 6522
          | |||||||||||  || ||||| |||  |||||| |||||| |||||| |||||
Db   2398 GCAAGGTCGGCCGATCATCCAGGGCCGTGGGAGATCACCAACTTCTACAAGGGCGACGG 2457

Qy   6523 GCTTCAAGGACGCCGGCAACCTGGGGCTCGCCCCCGCCCCCGGTTCCACCGGCAAGG 6580
          |||||||||        ||||||| ||  |||  |||||||||||||||||||
Db   2458 CGTTCAAGGACCAAGAACAACCTCGGCATCGGCCGTCCCGGCCCGGTTCCACCGGCAAGG 2517

Qy   6581 -GGGCTCTCCCCAGGGCGGTGAACCTCCGGTGTCGACGGGGTCGAA------GAACC 6633
           |||| |||   ||||||| |||||| ||  ||| |||||||||||
Db   2518 CGGGGCGCCCGCCCGACCAGGCGGCCACCAACCTCTCGGTTCTACGGCGACGCCGCCACC 2577

Qy   6634 TCGACGCCTCCTACGCTTCGTGAAGTACATGAGCTCCGCCCAAGGTGCAGCAGACCA 6693
          || ||||||  || ||  ||||| ||||||||  ||||| || |||||||| |||||
Db   2578 AGGAGGCGGGCGCAGAGCGGTTCGTGAACTTCGTGAACTTCATTGACCGTCCCGGAGAGCAACCGTCG 2637

Qy   6694 CCGAGAAGCTGAGCCTGCTGCCTGCACACGTCCTCGTCCTACGAGGTCCGTCGCGG 6753
          || |||     ||||| ||||||||||||  |||||| | ||||| ||||||||
Db   2638 CCCTGAAGAACTCCAGCGTGCCGACCGTCGCACGACGCGCTAC---ACCGCGAAGGTCAAGG 2694

Qy   6754 ACAACAGAGATGTGAAGTTCTTCAAGCCCGGTCGAAGGGCGTCGACAAGGCCGTCGAACCGGCCGTGGA 6813
          |||   |||     || |||||||||||   |||||   ||| |||  |||||||
Db   2695 CCGAAGGGCGGGTCATCGGGCTTCCAGACGTTCCAGACGCGCTTCCAGAGCAGCCGCCCGGCGC 2754

Qy   6814 TCGCCGAGGGGCAATGCCCTTCCTCGAGCCGAT 6844
          |  ||||| | ||||||||||||||||||
Db   2755 TGCCCGGAGTACAGCTCCTCGTGGACGCCGCGCT 2785
```

RESULT 12
SC10B7/c

LOCUS       SC10B7              33517 bp    DNA      linear    BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid 10B7.
ACCESSION   AL355752 AL645882
VERSION     AL355752.2  GI:20520696
KEYWORDS    acetyltransferase; alpha-glucosidase; arsC; bi-functional protein
            (secreted alpha-amylase/dextrinase); carboxylesterase; glnE; glnII;

```
               integral membrane protein; iron-sulfur oxidoreductase; lipoprotein;
               malE; malF; malG; malR: reductase; regulatory protein; secreted
               protein; tetR-family transcriptional regulator; transcriptional
               regulator; two-component system response regulator; two-component
               system sensor kinase.
SOURCE       Streptomyces coelicolor A3(2).
  ORGANISM   Streptomyces coelicolor A3(2)
               Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
               Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE    1  (bases 1 to 33517)
  AUTHORS    Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
               Kinashi,H. and Hopwood,D.A.
  TITLE      A set of ordered cosmids and a detailed genetic and physical map
               for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL    Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE    97000351
   PUBMED    8843436
REFERENCE    2  (bases 1 to 33517)
  AUTHORS    Seeger,K.J. and Harris,D.
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 33517)
  AUTHORS    Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-MAY-2000) Streptomyces coelicolor sequencing project,
               Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
               CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
               David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
               Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT      On May 9, 2002 this sequence version replaced gi:7799230.
             Notes:
             Streptomyces coelicolor sequencing at The Sanger Centre is funded
             by the BBSRC and Beowulf Genomics
             Details of S. coelicolor sequencing at the Sanger Centre are
             available on the World Wide Web.
             (URL, http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
             numbered using the following system eg SC7B7.01c. SC (S.
             coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
             strand).
             The more significant matches with motifs in the PROSITE database
             are also included but some of these may be fortuitous. The length
             in codons is given for each CDS.
             Usually the highest scoring match found by fasta -o is given for
             CDS which show significant similarity to other CDS in the database.
             The position of possible ribosome binding site sequences are given
             where these have been used to deduce the initiation codon. Gene
             prediction is based on positional base preference in codons using a
             specially developed Hidden Markov Model (Krogh et al., Nucleic
             Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
             of Bibb et al., Gene 30:157-66(1984) as implemented at
             http://www.nih.go.jp/
             jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
             correct initiation codon. Where possible we choose an initiation
             codon (atg, gtg, ttg or (att) which is preceded by an upstream
             ribosome binding site sequence (optimally 5-13bp before the
             initiation codon). If this cannot be identified we choose the most
             upstream initiation codon.
             IMPORTANT: This sequence MAY NOT be the entire insert of the
             sequenced clone. It may be shorter because we only sequence
             overlapping sections once, or longer, because we arrange for a
             small overlap between neighbouring submissions. Cosmid 10B7
             Overlaps cosmid 3H12 on the AseI-C genomic restriction fragment.
FEATURES             Location/Qualifiers
     source          1..33517.
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 10B7"
     RBS             462..467
     gene            473..1168
                     /gene="SC10B7.02"
                     /note="SCO2207"
     CDS             473..1168
                     /gene="SC10B7.02"
```

```
                     /note="SC10B7.02, hypothetical secreted protein, len: 231
                     aa; identical to previously sequenced TR:Q9X955
                     (EMBL:Y13833) Streptomyces coelicolor hypothetical 23.4 kD
                     protein precursor. Contains a possible N-terminal region
                     signal peptide sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical secreted protein"
                     /protein_id="CAB90855.1"
                     /db_xref="GI:7799232"
                     /db_xref="SPTREMBL:Q9X955"
                     /translation="MSRRRTLGTQKKIVLLVGAAVAGGGAFLAATASNAAQTTPEAR
                     TKAAQDSGVCQGLATALGNNQRFIEGQRADPDAQSQARIENREAVVAEIERKQAASGC
                     AVGESAQDPQAAQPPQSGQDTGGQAGDQAGDQAGNQTGDRAGGAAQAGEQVCAGSTVT
                     LSGEGGAPASSDRFPVGTKLRVTNLDNDKSTTVEVNSVSGSCALLNNAAFEQVREEG
                     KFLIRRALIEKVG"
     gene            complement(1171..2043)
                     /gene="SC10B7.03c"
                     /note="SCO2208"
     CDS             complement(1171..2043)
                     /gene="SC10B7.03c"
                     /note="SC10B7.03c, probable carboxylesterase, len: 290 aa;
                     identical to previously sequenced TR:Q9X956 (EMBL:Y13833)
                     Streptomyces coelicolor hypothetical 30.8 kD protein, 290
                     aa and similar to TR:Q59248 (EMBL:X82892) Bacillus
                     subtilis carboxylesterase NP (EC 3.1.1.1) Nap, 300 aa;
                     fasta scores: opt: 438 z-score: 516.3 E(): 2.2e-21; 31.5%
                     identity in 292 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative carboxylesterase"
                     /protein_id="CAB90856.1"
                     /db_xref="GI:7799233"
                     /db_xref="SPTREMBL:Q9X956"
                     /translation="MSARGPYDAGRFHDAYDKVMAKWPAGTESVTVPTPGATHVHLT
                     GPADGRPLVLLPGGGAATSACWYAQAAELSRTHRVHAVDLIGAPGRSTPAGDRHPRTV
                     ADLAGWLDALLDGLGIAETDLGHSYGAWIALHHTLHAPGRVRRLFLLDPTQCFAGFK
                     AAYLLRALPMLLRPTPRRVRAFLEWETGATPLDADWVALQEAAVGFPERRPVTGPRPA
                     PEALRALEVPLLLAGNSRTHDAAEVAARARTLLPHVETEALADVSHHALPQSAPPG
                     LGRRLGDFLTAPAG"
     RBS             2142..2147
     gene            2155..2679
                     /gene="SC10B7.04"
                     /note="SCO2209"
     CDS             2155..2679
                     /gene="SC10B7.04"
                     /note="SC10B7.04 probable transcriptional regulator, len:
                     174 aa; identical to previously sequenced TR:Q9X957
                     (EMBL:Y13833) Streptomyces coelicolor hypothetical 19.0 kD
                     protein, 174 aa and similar to TR:Q9XA85 (EMBL:AL096837)
                     Streptomyces coelicolor putative marR-family
                     transcriptional regulator, SCF43A.30c, 168 aa; fasta
                     scores: opt: 276 z-score: 336.5 E(): 2.3e-11; 39.9%
                     identity in 153 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="CAB90857.1"
                     /db_xref="GI:7799234"
                     /db_xref="SPTREMBL:Q9X957"
                     /translation="MPEEMAEPEETEGSTRMSDEPEELEIVHLLRAVTVELGAHSARF
                     AQRNGMHPTDVRALIALMDAARAGEATTAGHLGAALGLNSAGTTALVDRLERAGHVRR
                     VRDERDRRVTVEDERAVALGWSHFGPLIGRAVELLRGYDERELAAIRGFLTGVREA
                     AADDGREPRHNEPG"
     RBS             2801..2805
     gene            2816..3847
                     /gene="SC10B7.05"
                     /note="SCO2210; glnII"
     CDS             2816..3847
                     /gene="SC10B7.05"
                     /note="SC10B7.05, glnII, glutamine synthetase, len: 343
                     aa; identical to previously sequenced TR:Q9X958
                     (EMBL:Y13833) Streptomyces coelicolor glutamine synthetase
```

(EC 6.3.1.2) (glutamate-ammonia ligase) GlnII, 343 aa and
highly similar to SW:GLN2_STRVR (EMBL:X52842) Streptomyces
viridochromogenes glutamine synthetase II (EC 6.3.1.2)
(glutamate-ammonia ligase II) GlnII, 343 aa; fasta scores:
opt: 2196 z-score: 2572.0 E(): 0; 91.0% identity in 343 aa
overlap. Contains Pfam match to entry PF00120 gln-synt,
Glutamine synthetase and two matches to Prosite entries
PS00180 Glutamine synthetase signature 1 and PS00181
Glutamine synthetase putative ATP-binding region
signature"
/codon_start=1
/transl_table=11
/product="glutamine synthetase"
/protein_id="CAB90858.1"
/db_xref="GI:7799235"
/db_xref="SPTREMBL:Q9X958"
/translation="MTFKAEYIWIDGTEPTAKLRSKTKIITRAPAGLDALPVWGFDGS
STNQAEGSSSDCVLKPVFSCPDFIRGGEDILVLCEVLDTDMFHPSNTRAALAELSER
FAAQEPVFGLEQEYTFFKCTRPLGFPEGGFPAAQGGYYCGVGSDEIFGRDVVBAHLEN
CLKAGLGISGINAEVMPGQWEFQVGPLAPLEVSDLWVARWLLYRTAEDFEVSATLDP
KPVKGDWNGAGAHINFSIKAMREGYDAIITAAESLGEGSKPMDHVKNYGAGIDDRLTG
LHETAPWNEYSYGVSDRGASVRIPWQVERDKGYIEDRRPNANVDPYVTRLLVDTCC
TALEKAGQV"

misc_feature   2816. .3793
               /gene="SC10B7.05"
               /note="Pfam match to entry PF00120 gln-synt, Glutamine
               synthetase. score 559.70, E-value 1e-177"

misc_feature   2936. .2989
               /gene="SC10B7.05"
               /note="PS00180 Glutamine synthetase signature 1"

misc_feature   3470. .3520
               /gene="SC10B7.05"
               /note="PS00181 Glutamine synthetase putative ATP-binding
               region signature"

RBS            3891. .3894
gene           3903. .4832
               /gene="SC10B7.06"
               /note="SCO2211"

CDS            3903. .4832
               /gene="SC10B7.06"
               /note="SC10B7.06, unknown, len: 309 aa"
               /codon_start=1
               /transl_table=11
               /product="hypothetical protein SC10B7.06"
               /protein_id="CAB90855.1"
               /db_xref="GI:7799236"
               /db_xref="SPTREMBL:Q9K226"
               /translation="MEIDGVLPLRVETENGQSHARIRAAELRELVLRIGHAGDHFLVV

Query Match     12.5%;  Score 855;  DB 1;  Length 33517;
Best Local Similarity  63.0%;  Pred. No. 6e-57;
Matches 1683;  Conservative    0;  Mismatches 800;  Indels 188;  Gaps  16;

Qy  4335  CCCCGGCCGGCACCCGCCCCGCCCGCCGGCTGGCCCCGGCCGGCCGGTCAGCGTCCGGCG  4394
          ||-|-||||  |||   ||||| |||  ||| |-||-| | |  |||||||||  ||||
Db  30234  CCTCAGGTGGAACCGCACCCGCCGCCGCCCGAGCGCCGGGGGAATCAG--GG  30177

Qy  4395  GCGCCGGCCGGCTGCCCCCGGCGCCCGGCGCCGGCGCCGCCCGCGGTGACCCCGGCATGAA  4454
          |||||||| ||| |-|-||| |   |-|||  |  || ||||||||||||||||||||||
Db  30176  GCGCCGGCCGGCTGCCCCCGGCGCCCGGCGCCGGCGCCGCCCGAACCCGCACCCAACTCCGGCATGAA  30117

Qy  4455  GACGAATTCGGTGCCGCGGCGGGGCGGCGTTCCGCTCATCTCCTCCAGCAGTCGCT  4514
          |||||||| ||| || | ||| |   ||||| || | || ||||||||||||| |||
Db  30116  CACGAACTCGGTGCGCGGCGGGCGGGCGGCGGCGGCGATCCTCCAGGAGCGTGCGCACCGC  30057

Qy  4515  GACCTGCCCATCGCGGCCTTGACGGGCTGTCTGATGGTGTGCCGGCGGTCGGTGAAGGC  4574
          |||| | |   |  |  |||  ||||| | || |||||| |||||| ||| ||||||||
Db  30056  CGCCTGCCCATCGCGGGACCCGCGCTTGCGGACCGGTCGCGTGAGCGGTGGGTCGGTGAAGGC  29997

Qy  4575  CATGAGGCCGAGTCGTCGAAGCCGACGACCAGATGCACCGGAACCGTGAGACCCCG  4634
          || |||||||||||| ||||||||||  | |||||||||||| ||| ||||| |||
Db  29996  GATCAGCGGCCGAGTCGTCGAAGCCGACGACCGAGCTCCTTGGGCACGTCCAGGCCCAG  29937

```
Db  28856  ACGTC-GCCCGACGGCAGCAGCAAGCGGTCGGCGTCGCGGCAGGAGCGGCGGACGGGCCC  28798
                |  |||      |||       ||||||||| ||||  ||  ||
Qy   5635  AGACTTCCGCCCGATTCACTCAGGGGAGTTCACAATGCGGCTGGCATTGCGG  5694
               ||    ||     |    ||||||||||||||||
Db  28797  GCACCCTGCACCACACGGGCTTTCACCCTCAAGGAGAACTCATGCGGCTGGCATAGCCGG  28738
                                              |||||||||||| ||||
Qy   5695  CCACCGGCTGTTCGCGCCATCAGCACGGGCATCGGCGTGTGGCGGGGGGACACG  5754
           |||||||||||| | ||||   || ||||  ||||||||||||||||| ||||| ||
Db  28737  CCACCGGCGCTGGTGGCCGTCCCTCGCCCTGGCCGGCACGGACGCGTCGCGGAGACGGCCACA  28678
                                                           |||||||||||||
Qy   5755  GCGGAAGCGGTACCGACGGCGGCGCACGGAGCTGTCGGGGACCGTCACCTTCGGGACA  5814
           ||| |||| ||    |||  ||| |  |||||| | |||||  ||||||||||
Db  28677  GCGACAGCGGAGTCGGGCGGAC----------CCGGTCCACCATCACTTACTGGGACA  28633
                                                   |||||||||||||||||

Qy   5815  CGTCCAACGAAGCCGAGAAGGCGACG---TACCAGCCCCTCCCGGAGGGCTTCGAGAAGG  5871
           ||||||||||      |  ||||||||  |||   |||||||||||||| ||||||||||
Db  28632  CCTCCAACGCGACCAACGAGGCCGCCGACCTACACAGGGCCGCTGGTCAAGGAGTTCGAGGCCG  28573
               |||||||||  ||| |||||||||||||||||||||||||||||||||||||||
Qy   5872  AGCACCCGGAAGGTCGACGTACGTGACGTAGCGACCGCGCAAGTCGGTCAAGCGCCAAGT  5931
           |||||||||  ||||       ||| | |||||   |||||||||||| |||||   ||||
Db  28572  CCAACAAGGGGCGTCAACGTCAACTTCGTCAACGTGCCCCTTCGACCCAGCGCGCAGAACAAGT  28513
               |||||||||       |||||||||||||||||| ||   ||||

Qy   5932  TCAAGAACGCCCGGCGGCGCCAACTCCGGTCCCCGGAGGTGATTGCGGTACGGAGGTCGCGT  5991
           || | |||||| || || |||||||||||||||||| | ||| |||| |||||||||||||
Db  28512  TCGCACACCGCCGGCAGGGCTCCAA---GGGCGCCCCCGACGTGCTGCGGCTCCGAGGTCGGCT  28456
                  ||| |||  |||||||  | |||||||||||||||||||  |

Qy   5992  GGTCGCGGGACTTCGCCAGCATCGGCTACCTGCCCCGCTCGACGCCCGGCACGCCCGGCCCTCG  6051
           |||  || ||||| |  | ||||| | |||||||||||||||||| ||||||||||||||||
Db  28455  GGACGCCCGGCCTTCGCCAAGAAGGGCTTCTTCGCGCGCCTGGCACGGCGCACCGAGGCCCCTCG  28396
                        |   |||||||||  ||||||||||| |  | || |||||||||||||

Qy   6052  ACG---ACGGGTCGGACAAGTTCCAGCCACCTGATCAGCAGGCAGTACGAGGGGCAAGACCT  6108
                   | ||| ||||||||||||| | ||||||||| | || | ||||||||||||||||
Db  28395  CCGAGCAGGACAAGTTCCAGCCCCAACCTGATCAGCCCCGGCTCTACGACCAAGGGCAAGACCT  28336
                       ||||||||||||||| ||| ||||||||||  ||||||||||||||||||||

Qy   6109  ACGCGGTCCCGCAGGTCGATGCATGCACCCTGGCGCTCTTCTACACAACAAGGAACTGCTGACGA  6168
           |||  || ||||| ||| || || ||| ||||||||||||  ||||||||||||||  ||||||
Db  28335  ACGCCCGGCGTCCGCTGGTCACCGACACCCCCGTCTTCGTTCGTCTACACAACAAGGAACTGTTCGAGA  28276
                |||  || || ||| | ||| ||||||| |  ||||||||   || ||||||||||||

Qy   6169  AGGCCGGCGGTCGAGGTGCCGGGGCCTCCTCCCTCCGCCGAGCTGAAGACGACGGCCCGCCGAGATCA  6228
           |||||||  || ||||| | ||||| ||| | |||| || | |  |||||||| |||||| | |
Db  28275  AGGCCGGCGGCGTCGAGGTGCCGGACGACCTGGAAGAGGCCCGCGAAGAGGCCCGCCACCGTCA  28216
              |||||||  ||||||| | |||   ||| | | || |   |||||||||||||||||||

Qy   6229  CCGAGAGAACCGGCGGGCGAGCCGGCCGGCCTCTACTGCGGGGCGACGACCGTACTTGGTTCCTGC  6288
           ||||| ||  |  ||| || || || | ||| |||   |  |||||||||||  |  | |  ||
Db  28215  AGGACAAGACACCGCG---TCGACGGCTACTGGGCGTCCACGGCCGGCCTACTACGCGCAGC  28159
           |||| |||| | |||      ||||| ||||||| || || |||||||| || |||||||||

Qy   6289  CCTACCTTCTACGGGGAGGGGCGGCGGCGACCTGGTCGACGAGAGAACAAGACCGTCACGGTCG  6348
           ||||| ||| |  ||| |||||||||| ||| ||| ||||||||||| |||||||||| |||||
Db  28158  CGTTCCGTGTACGGCGAGGGCCGCCGACGCTCGACGCCGGCCGCCAAGAAGGTCACCGTCA  28099
              | ||| || |||| ||||||| || || || ||| |||  || |||||||||| ||||

Qy   6349  ACGACGAAGCCGGCGTGTGCGCGCCTACCGCTCATCAAGGACTTCGTGACAGCAGGCGG  6408
           ||||| |||||||| | |  |||||||||| ||   ||||||  |||||||| ||||||
Db  28098  ACGGCCGCGGCGGCGCAAGAGGGCTACGGCGCCGGCTGAGCCCTTTCGACGGCCAAGGGCC  28039
              ||| |||||||| | |  ||||||| || ||  |  | |||||| ||| ||||||||

Qy   6409  CCATCACCGACG------CGTCCGACGCGGCTGGAACAACATGCAGAAACGCCTTCAAGTCGG  6462
           |  || ||||||        ||||  ||| |||||||||||| |  |||||||||||| |||
Db  28038  TGCACAAGGCCGGCAACAACCTCGGCATCGCCGACCTCCACCATCCAGGAGCGCTTCGTCAGCG  27979
              |||| |||||||||| ||  |||| ||||| ||||||| ||  ||||||||||||| |||

Qy   6463  GCAAGGTCGCCATGATGGTCAACGGCCCTGGGCCATCGAGGACGTCAAGGCGGGAGCCC  6522
           |||||||||  |  || ||||||  || ||  |||  |||||||||||||||| |||||
Db  27978  GCAAGGTCGCGATCATCCAGGAGATCACCAACTTCTACAAGGGGCACGG  27919
           |||||||||| || || ||||| |||  ||  |  ||||| ||| |||

Qy   6523  GCTTCAAGGACGCCCGGCAACCTGGGGTCGCCCGTCCCGGCCCAGTGCCCGGACA--  6580
              ||||||||| | |  ||||||  |||  |||  |||||||||| || |||||||
Db  27918  CGTTCAAGGACAAGAACAACCTCGGCATCGCCGGCGTTCCACCCGGCAAGG  27859
           |||||||||| |  |||| ||| || ||||| | | ||||   ||| |||

Qy   6581  -GGCTCTCCCCAGGGGGGTGGAACCTCTCGGTGTACGGCGGCTCGAA------GAACC  6633
            |||||  |  |||||||| ||  ||   ||| | ||||| | | |||      ||||
Db  27858  CGGGGCGCCCCCGACCCGGCCGGTCGTTCGGACGACGCCGCACC  27799
            ||| | |||||||| |||||| ||||  || ||| |  |  |||

Qy   6634  TCGACGGCCTCCTACGCCTTCGTGAAGTACATGAGCTCCCGCCAAGTGCAGCAGAGACCA  6693
            ||| |||||||  |  ||| |   || | | || | ||||| ||||||||||| | || ||
```

```
Db  27798  AGGAGGCGGCGCTGAAGTTCGTGAACTTCATGACCTCGGCGGAAGTCCAGGAGGACCGTCG  27739
              |||      |||       ||||  |     |||  ||||  ||| |||
Qy   6694  CCGAGAAGCTGAGCCTGCTGCCACCGCACGTCCGTCCTCCTACGAGGTCCGTCGCGG  6753
                |||  |||   |   || ||||  ||||||||||||| ||
Db  27738  CCCTGAAGAACTCCACGCTGCCGACCGCGACGAGCCCTAC---ACCGCGAAGGTCAAGG  27682
              || ||||||| | ||||||||  || |  |  | ||||     |||   |||||||||
Qy   6754  ACAACGAGATGGTGAAGTTCTTCAAGCCGGTCGACAAGGCCGTCGACAAGGCCGTGGA  6813
           | |||||| |||||||| | |||||||||||| |||| | ||||||| || | |||
Db  27681  CCGACCCGGGCATCGGGCGGGTTCCAGACGGTCCTCCAGCCGCGCCCCGGCGC  27622
           ||||||||| ||  || || |  |||| ||| | | || | | | |||| |

Qy   6814  TCGCCGAGGCGAATGCCCTTCGAGCCGAT  6844
              ||   || | | ||||||||||||||
Db  27621  TGCCCGAGTACAGCTCCCTGTGGACGCCCGCT  27591
            | |||||| | || ||| || |  || || ||
```

```
RESULT 13
SVI011500
LOCUS          SVI011500             39250 bp    DNA     linear   BCT 24-MAR-1999
DEFINITION     Streptomyces violaceoruber Tu22 granaticin biosynthetic gene
               cluster.
ACCESSION      AJ011500
VERSION        AJ011500.1  GI:4218538
KEYWORDS       acyl carrier protein; aromatase;
               CDP-4-keto-6-deoxyglucose-3-dehydratase(E1); cyclase;
               cyclase-dehydratase; disulphide bond-forming protein;
               dTDP-1-glucose synthase; dTDP-glucose-4,6-dehydratase;
               dTDP4-keto-6-deoxyhexose reductase; FMN-dependent monooxygenase;
               FMN-NADH oxidoreductase; glycosyl transferase; gra-orf 1-37 genes;
               hydroxylase; pathway-specific transcriptional activator; polyketide
               ketoreductase; polyketide synthase chain length factor; polyketide
               synthase keto-acyl synthase; serine-threonine protein kinase;
               transcriptional activator; transmembrane protein; tRNA-Ala gene;
               tRNA-Alanine; two-component response regulator; two-component
               sensory kinase.
SOURCE         Streptomyces violaceoruber
  ORGANISM     Streptomyces violaceoruber
               Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
               Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE      1  (bases 1 to 39250)
  AUTHORS      Sherman,D.H., Malpartida,F., Bibb,M.J., Kieser,H.M., Bibb,M.J. and
               Hopwood,D.A.
  TITLE        Structure and deduced function of the granaticin-producing
               polyketide synthase gene cluster of Streptomyces violaceoruber Tu22
  JOURNAL      EMBO J. 8 (9), 2717-2725 (1989)
  MEDLINE      90060034
  PUBMED       2583128
REFERENCE      2  (bases 1 to 39250)
  AUTHORS      Bechthold,A., Sohng,J.K., Smith,T.M., Chu,X. and Floss,H.G.
  TITLE        Identification of Streptomyces violaceoruber Tu22 genes involved in
               the biosynthesis of granaticin
  JOURNAL      Mol. Gen. Genet. 248 (5), 610-620 (1995)
  MEDLINE      96027933
  PUBMED       7476861
REFERENCE      3  (bases 1 to 39250)
  AUTHORS      Ichinose,K., Bedford,D.J., Tornus,D., Bechthold,A., Bibb,M.J.,
               Revill,W.P., Floss,H.G. and Hopwood,D.A.
  TITLE        The granaticin biosynthetic gene cluster of Streptomyces
               violaceoruber Tu22: sequence analysis and expression in a
               heterologous host
  JOURNAL      Chem. Biol. 5 (11), 647-659 (1998)
  MEDLINE      99051446
  PUBMED       9831526
REFERENCE      4  (bases 1 to 39250)
  AUTHORS      Hopwood,D.A.
  TITLE        Direct Submission
  JOURNAL      Submitted (22-SEP-1998) Hopwood D.A., Genetics, John Innes Centre,
               Norwich Research Park.Colney Norwich, Norfolk NR4 7UH, U.K
FEATURES             Location/Qualifiers
     source          1..39250
                     /organism="Streptomyces violaceoruber"
                     /strain="tu22"
                     /db_xref="taxon:1935"
     gene            complement(1099..2417)
```

```
CDS          /gene="gra-orf7"
             complement(1099..2400)
             /gene="gra-orf7"
             /codon_start=1
             /transl_table=11
             /product="putative serine-threonine protein kinase"
             /protein_id="CAA09628.1"
             /db_xref="GI:4218539"
             /translation="MGEVPAGRYELIDPIGRGGVGAVWRAWDHRRRYVAAKVLQQSD
             AHTLLRFVREQALRIDHPHVLAPASWAADDDKVLFTMDLVAGGSLAHVIGDYGPLPPR
             FVCTLLDQLLSGLAAVHAEGVVHRDIKPANILLDVTGRPHLRLSDFGISMRKGEPR
             LTETNYVVGTPGYFAPEQMGAEPDFPADLFAVGLVALYLLQGRKPDSRALIEYFLAH
             GTPGAPQGIPEPLWQVLAGLLQPDDQARFKTATGVRKALTAAVEMLPEAGPGEEPVEV
             FDQIGPCPPASARTARAPSPAAPAPVRVQRHPEPQOPQOPLPTLSETGSFHLPPPPR
             QOPAPPPQPVAAPAPHPAAVTPLAHTDPALAVPTAAVPYEQASTRSYTAHAPRIPA
             QASPAAQSPAQRPGPPPKVAIPVLFVALICFAVGIWALIQT"

RBS          complement(2414..2417)
             /gene="gra-orf7"

gene         complement(2474..3031)
             /gene="gra-orf8"

CDS          complement(2474..3022)
             /gene="gra-orf8"
             /function="unknown"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA09629.1"
             /db_xref="GI:4218540"
             /translation="MDAAQQEATARARELQRSWYGEPLGALFRRLIDDLGLNQARLAA
             VLGLSAPMLSQLMSGQRAKIGNPAVVQRVQALQELASQVADGSVSAAEATDRMDEIKK
             SQGGSVLTGTSQTNSSGAPTVRRVVREIQSLLRSVAAGHIIDAADSLAPAHPELAE
             FLRVYGAGRTGDAVAHYESHQN"

RBS          complement(3028..3031)
             /gene="gra-orf8"

gene         3152..3226
             /gene="tRNA-ala"

tRNA         3152..3226
             /gene="tRNA-ala"
             /product="tRNA-Ala"

gene         3232..3267
             /gene="tRNA ala"

terminator   3232..3267
             /gene="tRNA ala"
             /note="putative terminator (stem 16 bp, loop 4 bp)"

gene         complement(3290..3346)
             /gene="gra-orf8"

protein_bind complement(3290..3346)
             /gene="gra-orf8"
             /note="5 CGA repeats"
             /bound_moiety="SARP"

gene         3454..4313
             /gene="gra-orf9"

RBS          3454..3457
             /gene="gra-orf9"

CDS          3462..4313
             /gene="gra-orf9"
             /codon_start=1
             /transl_table=11
             /product="pathway-specific transcriptional activator"
             /protein_id="CAA09630.1"
             /db_xref="GI:4218541"
             /translation="MDVGILGPLAVRLNGQSIVPSAGKPRQLALLALAIRAERVVTVPT
             LMEEIWGDAIPKSAATTLQTYILHLRRKVTAALRAGOPGAAADRAGTGAGADAKSVL
             STCFGGYGVRLALREFQSLAARGSVALEAGDARAASAMLERALGLWRGTALVD
             VPVGRVLGVEVLGMEEQRTRVLEQRIEADLLGRHASLLAELRMLVARHPMNENLAAQ
             LMTALYRSGGVWRALEVFQELRRTLIEELGVEPSHRLQLLHRAVLSGELELDPRHAGL
             ELMRTCG"

gene         4381..4413
             /gene="gra-orf9/gra-orf10"

terminator   4381..4413
             /gene="gra-orf9/gra-orf10"
             /note="putative bidirectional terminator (stem 12 bp, loop
             9 bp)"

gene         complement(join(4418..5038,5051..5140))


CDS          /gene="gra-orf10"
             complement(4418..5026)
             /gene="gra-orf10"
             /codon_start=1
             /transl_table=11
             /product="putative two-component response regulator"
             /protein_id="CAA09631.1"
             /db_xref="GI:4218542"
             /translation="MIRILLAEDMEMVRGALVALLGLEDDLEVVAQVERGDKILPEAL
             KVRPDVAIIDIGLPGIDGIKAAALLAEELPSCRTLILTSLSRPGVLRRALDAKVAGFL
             PKDASPQDLAEAVRRIAAGHRAIDPQLALAAWDSAESPLTEREMEVLRLTAEGDEAPE
             IAARLHLSTGTVRNYLTSVVTKLNARNRVDAVRIAHESGWLV"

RBS          complement(5036..5038)
             /gene="gra-orf10"

protein_bind complement(5051..5140)
             /gene="gra-orf10"
             /note="4 CC/AA repeats"
             /bound_moiety="SARP"

gene         complement(join(5125..6345,6567..6602))
             /gene="gra-orf11"

CDS          complement(5125..6345)
             /gene="gra-orf11"
             /codon_start=1
             /transl_table=11
             /product="putative two-component sensory kinase"
             /protein_id="CAA09632.1"
             /db_xref="GI:4218543"
             /translation="MNSREQQFDELDSSASYLSHEDSDLPAPRLARTVLLAAMAGYSL
             VTVVNLLGAGLSPAETCVGFTALLVVFGLQOVAHSRPQARGSPYWRRVTTLTLQAVFCF
             VPLALFHMAWGAMGGFLAGSVLLLPGRAAWPLYGAIGASLFAYSAAGHTLLDSVYT
             LEATLVTGLVVFGLSSLTGLVTRLHAHRADLARLAVAGERLRFSRDLHDLLGYNLSAI
             TLKTELVRRLVPDMPERARDETDSILALSRQSLADVRKVATGYRDMSLREEAAATRSM
             LEAAGIEVEAAIDIGPLPSQIDTVLATVLREGVTNALRHSRVHHCTIRCWEKDGVVTL
             SLGNDGVDPEDGGYEGVDRDARRADDPGGTGLGNLTYRLRTRVGGSLTTELSDNGWFRM
             TARAPVHPHVCTTR"

protein_bind complement(6567..6602)
             /gene="gra-orf11"
             /note="2 CC/GA repeats"
             /bound_moiety="SARP"

gene         complement(6688..8981)
             /gene="gra-orf12"

terminator   complement(6688..6711)
             /gene="gra-orf12"
             /note="putative terminator (stem 8 bp, loop 8 bp)"

CDS          complement(6716..8971)
             /gene="gra-orf12"
             /function="unknown"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA09633.1"
             /db_xref="GI:4218544"
             /translation="MTGEAVYLVSGRGVRAVRGAPAEALAPLLDGTRTPDEVLREAAR
             ARDARRPLPRGVRRALAELWSAGLLRAHPDPAATPGTGAGPPAAHDRSAEAFWDLAGT
             TPPARTAVRLLVLDGPVSEEEARAACAASGLDVGRPGAQASDAEVTLVLCTDYLDPRL
             ADVDALQRAAGRPWILARPGGPEAWTGPVLRPDDGPCWHCLAERLRLQRGSEAPLHRA
             LGCTSPVARPDASLAASRAAGLHTAVLETAKWLAGVPGTRDAVHTVDALITLDSARHP
             VARRPQCPSCGDAGLVRARVRAPFVPAPRPRKAADGGNGHRALSPREVLERYGHLVGPV
             TGVVEVRRDRPSPAFTHSFLSGHNLAMRAPTLAGHSAGLRALSGGKGLTSEEARVSA
             LCEAVERYSGSRHGDEPVVRDSFRAFGPAAVHPDTVRLYDPRQLRDRDAWNASGSPFQ
             YVGRPFDEDAPVDWTPVWSLTGCVGVQRLLPTSLLYFDADPAAPLRADSNGNAAGSSPED
             AFVQGFELVERDAVALWYNRTRAPACDLDAFAASGDAWIEPLRAGYRSLGREIWAL
             DLTSDFGIPVVAALSRRTDGPPEDVLFGAHLDPRIAVRALTEMGQLPAVVDARP
             DSGGYTVDDPHALSWWRTATVAGQPHLLPDPQRAPSRPGSWTYTARADLSEDVAALAA
             LADERGLEVLVLDQTRPDIGLPVKVVVPGMKQFWPRFAPGRLFDVVALGRLAEPTP
             YERLNPLPLFV"

RBS          complement(8978..8981)
             /gene="gra-orf12"

gene         complement(9059..9245)
             /gene="gra-orf13"

CDS          complement(9059..9232)
             /gene="gra-orf13"

Query Match   12.4%;  Score 850.8;  DB 1;  Length 39250;
```

```
Best Local Similarity  65.1%;  Pred. No. 1.2e-56;
Matches 1327;  Conservative  0;  Mismatches  687;  Indels  24;  Gaps  4;

Qy   2269 TGAAGGGCCTGGTCCTGCCAGGTGGAACCGGCAGCAGACTGAGGCGTTCACCCACACCG 2328
          |||||| ||||||  |  |||| ||||||||| |||  ||||||   |||  ||
Db  12913 TGAAGGGCCTGATTCTGTCCGGAGGGGATGGGGACCAGGTTGCGTCCCTTCACCTATTCCA 12972

Qy   2329 CCGCCAAGCAGCTGCTCCCCATGCCGCCAACAAGCCGTGCTGTCTTCTACCGCGTGGAGTCCC 2388
          |||| |||||||||||| |||||  || |||||| |  |  |||||   |||| ||||||||
Db  12973 TGCCCAAACAACTGGTCCCCGGTCGCGCATCCTCCGTGCACTGCGTCTGGAACG 13032

Qy   2389 TCGCCCCGGGGTGTCCGGAGGCCGGCGCTCGTCGTGTGGGCGTACGGCCGGGAGATCC 2448
          ||  |||||| |||||||| |||||||| ||||| ||| |||  ||| |||
Db  13033 TCCGTGCCATCGGTGTCGAGGAGTCGGCCGTCGTCGGGGACCGGGGACGAGGTCC 13092

Qy   2449 GGCAACTCACCGGACGACGGCACCGGTTCGGGTTACGGCGTACCACCTCCACCAGCCCC 2508
          |  |||| |||   || |    ||| |||  |  |  ||| ||||||  |||| |||||
Db  13093 GGGCGGTGTCGGTGACGGTCCGGCGCTCCGCGTTCGGCGGCTCACCGGTGTCCAGCAGGAGG 13152

Qy   2509 GCCGAGTCGGTCTCGCGCACGCGGTGCGCATCGCCCCGGCGCTTCCTGGCGACGACGACT 2568
          ||||||||| |||||| |||||||||||  |||| |||| |||||   |||||||||||
Db  13153 CGCCGAGTCCCACTGCGCTCTCCATCGCCGAGGAGTTCCTGGTGACGAGGACT 13212

Qy   2569 TCCTGCTGTACCTGGGGACAACTACCTGCGGGCGTGCCCCAGGGCGTCACCGACTTCGCCGCCAAT 2628
          || |||||| ||| |||||||    |||||| ||||   ||| |||| ||||||||||
Db  13213 TCGTGATGTAACCTGGGGGCACAACATCTTCGGCTGAGGCGATCGCGCCGAGGCCCGCGCGT 13272

Qy   2629 CGGCCCGATCCGCGGCCGGCGTGCTGCTCACCCCGGTCGCGGACCCGTCCCGCCT 2688
          ||||||||  ||||||||||||| ||||||  ||| |||  ||||||||||
Db  13273 TCCGGACGAACGTTCCGCGCGCTGCTGCTCCAAGGTGGCCCGGACCCGCTGCCT 13332

Qy   2689 TCGGCGGTCGCGGAGGTCGACGCGGACGGGAAGGCTGCGTTGGAGGAGAAACCGGACG 2748
          ||||||||  |||  ||| |   ||||  ||| |||||| ||| ||| |||||||
Db  13333 ACGGCGTCGCCGAGACGGCGATGCCACCCGGTCCTCAGGGTCGGTGGACGCCGAGC 13392

Qy   2749 TCCTGCTCAGCTCGCCCCCAGGGGACGACTACCTGCGCTCGCGTACGGCGTGTACGACCTTCAGCGAGG 2808
          ||||||  |||||| |  ||||||||| || |||||| ||  ||||||||| |  |||| |||||||
Db  13393 GGCGCAGCGCGAGGACGAACTGCGGCGCGGCGCCGGGCCGGCGCGCTCGTCTCCACGACG 13452

Qy   2809 CGGTACGGCGGCCATCACCCCCTCCGCGGCGAGCGTGGAGATCACCCACGCCGTGCAGT 2868
          | ||||||  | |||||| |  | ||||||||| |   ||| || ||| ||||||||
Db  13453 CCGTCCGGGCGATCGAGCGCCGGAGCCGGCGGAGCTCGAGATCACCGAGGCCGATCCAGT 13512

Qy   2869 GGATGATCGACACCGGGGGCTCGCGCGTACGGCGGCCCTGGCCCTGGCGGCCTGGCGGCGACA 2928
          |||||| ||||||  |  ||| ||| |    | |||| |||||  ||||||||
Db  13513 ACCTGTGGAGCGCGCGAGGACGATGCCGCCCGCGGGCTCGGTGTCGACCACACTG 13572

Qy   2929 CCGGCAGCGCGGAGGACGAATGCTGGAGGTCGAGACTGTCACCGGACTGGAGGGCC 2988
          ||||||  ||||||||||| | |||  |||||  |||||||
Db  13573 CCGGCAGCGCGGACGACTGGTCGCAATGGCGGTCCGGTACTCCGTCGGGCGCTCCGCCCCCG 13632

Qy   2989 GCATCGAGGGGAAGTCGACGCGCCACAGCACCGTCGACGCCGGGTCCGGGTGGCCGAAG 3048
          || |||| |  ||  |||||| ||  || ||||  ||||  | |||
Db  13633 GGTACACGGCCGGAGTCGACGGCCAGTGGCTGCGGGACCGTGGTCGTCGAGGGGG 13692

Qy   3049 GCCCGGATCGTGCCGGGGTCACACGTGGCCGGGTCTGGTGATCGGCGGCGGGGTGCCGTCG 3108
          |||||| | |||||||  ||  || | ||||||||
Db  13693 GACGGTGGTCCGTGGAGCGACTCGGTCGGACCGGCTCGGGCGGCCAGCCGTCG 13752

Qy   3109 TCAGCAACTCCAGTGTCCGGGGTACACCTCGGGCGCGGTCGAATGCGTACGGGACTACACCGCACT 3168
          || |||||   | ||| |||    | ||| || ||||||  ||||||||
Db  13753 TGCGGCGCACCGGAGGTCGAGGCGAGCTCGTACGGGTACCGGGACGACG 13812

Qy   3169 GGCCGACACCCGGAAGTGGTACGACGGGCACGAGGAGCGTCCCCATCG 4300
          ||   |||||| |||| |||||||||| | |||||  |||||| |||
Db  13813 GGGGGCGGTCGGGTGGACTGGTACCGGGGTACCGGCACCACCCCGACTTCTGCGCGGCGGTCC 13872

Qy   3229 AGGCGTCCCTCATCGGGCGCGGGCCGCGTCGCGGCGGGGCCGGCGGCGTCCCCCAGGCTC 3288
          |||  ||||  ||||||  |||||  ||  |||  |||||||
Db  13873 CCGGCTCGCCTCATCGGGCGGCGGCTCGCGCCGGCGCCGT------GCGCACCGGCGAGGCGGCCGAC 13926

Qy   3289 ACCGACTGGTGATCGGCGACCAGCAAGGTGTATCTCACCCCATGACCACGACCATCCT 3348
          ||   |||||| |  |  ||||||    |||   |||| ||  |
```

```
Db  13927 GGCGGCTCATCATCGGGACGACCATACGCAGGCGGACGGCGAGGTGGCGGCAT------GAGGCTCCT 13980
Qy   3349                                                                      3404
Db  13981 GGTCACCGGCGGCGGCGCCCCGGCGGGCTTCATCGGTTCATCGGCTCGCTCGCGCACTACGTGCGGGGGAG 14040

Qy   3405 --CCGGGGCCCCCGGCGGGCGGGGGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCCGCAG 3462
          ||||||||||| |||||||| ||| |||||| ||  |||||| ||  |||||||||
Db  14041 CTACCCGGAGTCCGACGACGTCGACGTCACGTCGTCGACTCACTCGCCACTCACGCGGGGGCG 14100

Qy   3463 CCTGCCGGGTGCAGCGGCGTGCGTGACCATCCGGCCTCACCTTCGTCAGGACGACGT 3522
          ||| |||||  |    |     |  |||||   || |||||||  ||||||||
Db  14101 CCGGCGAACGGCTGCCACTTCGTCGTCACGGCGCACGGCGACAT 14154

Qy   3523 GTGGACACCCGCTCGTCGACACGCTGGCCGCGACGCAGCCGGCTGCTTCGC 3582
          ||||||| |  ||   |  ||| ||||||  |||| | ||||| | ||| |||
Db  14155 CTGGCCACCGGCGACCTCCCACGGCGGCGCCTCCACGCGGGGGTCGTGCACTTCGC 14214

Qy   3583 GGCCAGTCGCACGTCGACCGCTCATCACCGACAGCCGGTGCCTTCACCCGCACCAACGT 3642
          ||||||| |||||| ||  ||| |||| ||||||| ||||| |||||||||
Db  14215 CGCCGAGTCGCCCACGTCGGACCGCTCCCGGCAGTTCGTCGTCCGTACGAACGT 14274

Qy   3643 GCTGGCGACCGTCGCAGCTGCTCGTGCCCACGTGTGCGCACCTTCGTCGTGCA 3702
          | |||||||||| ||||||| |||||||  ||||||| |||||||| | |||||||
Db  14275 CATGGGCCACCAGAGTCGCTGCTCACGCGGTCCACGACCTGTGGACCGAGTCTGCA 14334

Qy   3703 CGTCTCCACGACGAGGTGTACGGCTCCCCTCCCCGGCGGGGAGGACCC 3762
          ||||||||||||||||| |||||||||| | ||| || | ||| | ||||
Db  14335 CGTCTCCACGACGAGGTGTACGGCTCCGGTGGTTGGACAGCGGACCGGACCACCGAGGACGCCC 14394

Qy   3763 CCTGCTTCCGACCTTCGCGGCGTGACGCGGCCATGGGCGTCGCTCGC 3822
          ||| || |||  |||||| |||||||  |||| |||  |||||||||
Db  14395 CCTGCGTGCCCAACTCGGTGTTCGACGGTCCAAGGGCGTCCAAGGCGTCCGCGCC 14454

Qy   3823 CCACCACCGGACCACCCGACGGCCGGCCCACTGCTCGTGTTCGAACAACTTCGGGCCC 3882
          ||||| | | |||||| ||| | |||| ||||| ||||  |  ||||||||||||
Db  14455 TACCACGTACGCCACGGACTCGACTCGCGCCATCACCCCGTGCTCCAACAACTACGGCCC 14514

Qy   3883 CCACCAGCATCCGAGAAGCTCATACCGCGTTCCTGACCAGCGCGCCGTTCGCGAC 3942
          ||||||  |||||||||||| || | ||| | ||||  || ||| ||||  |||
Db  14515 GCGCCAGCGCGGAGAAGCTCATCCCGAACTTCGTCAACCCGGTGACCGGTCCGGGTCGGCGA 14574

Qy   3943 CGTTCCCCTTACGGCCGACGGCGCCCTGGCCCCTGGCGCACGTCGGCGACCACGTG 4002
          || ||||| | |||  | | ||| || ||||  ||  | || |||  ||||
Db  14575 GGTTCGGTCGTGTACGGGCGCGACGACGCAACGTGCGCCCGGTGGCGCCGGACCACTG 14634

Qy   4003 CAGGGCCGTCGAACTCGTCCCGTGTCGGACTGGAGGGCC 4062
          || || |||||||||| | | ||| |||||||||||||
Db  14635 CCGGGGCGCTGAACCTGGTGTCAACAAGGGCCGGAGATCCGGCCGAGATCTACAACATCGGCGG 14694

Qy   4063 CGGCACCTCGCTGCCCAACCTGGACGCCGGTTGCTCGGCGCGTCGGGGCGGG 4122
          |||  ||||||  ||||| ||||  |||| | |||| || ||  |||
Db  14695 TGGCAGCGGTGATGAGCAACCGGAGATGAACGAGGACCCTGCCGGCGGGACCGGGGCGGA 14754

Qy   4123 CCGGAGCGCGCATCGTCCACGTCCAGGGCGTCGAGAACCGCGAAGGGCTACGGCGTCGA 4182
          |||  || || ||||| | |||||||| | |   | |||| |||||||||
Db  14755 CTGGGACACGTGCCGGGCCATGTGCCGCCGGCACGGCACCGCTCCGCTATGCGATCGA 14814

Qy   4183 CCACAGCAAGAATCACCGGACGTTACCGGCGACGACGCGCT 4242
          |||| ||| ||| |||||||| || ||||| ||| ||| |||
Db  14815 CGACTTCCAGATCCCGCGGAGAGCTCGGGGCGGTACGCGGGCGCCGACTTCGCGGCGGCT 14874

Qy   4243 GGCCGACACCCGCAAGTGGTACGACGGGCACGAGGAGCGTCCCCGG 4300
          |||||||| ||| ||||||||||| || | ||||| | ||  |||
Db  14875 GGGGGCGGTCGGTCGTGGGGTACCGGTACCGGGCCGACCACTTCGGCGCGGCGGTCCTCC 14932
```

RESULT 14
AF164960

```
LOCUS        AF164960                8018 bp    DNA     linear   BCT 07-FEB-2000
DEFINITION   Streptomyces fradiae strain T#2717 urdamycin biosynthetic gene
             cluster, partial sequence.
ACCESSION    AF164960
```

```
VERSION     AF164960.1  GI:6002927
KEYWORDS    .
SOURCE      Streptomyces fradiae
ORGANISM    Streptomyces fradiae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 8018)
AUTHORS     Faust,B., Hoffmeister,D., Weitnauer,G., Westrich,L., Haag,S.,
            Schneider,P., Decker,H., Kunzel,E., Rohr,J. and Bechthold,A.
TITLE       Two new tailoring enzymes, a glycosyltransferase and an oxygenase,
            involved in biosynthesis of the angucycline antibiotic urdamycin A
            in Streptomyces fradiae Tu2717
JOURNAL     Microbiology 146 (Pt 1), 147-154 (2000)
MEDLINE     20121748
PUBMED      10658661
REFERENCE   2  (bases 1 to 8018)
AUTHORS     Faust,B., Westrich,L. and Bechthold,A.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUL-1999) Pharmazeutische Biologie, Universitaet
            Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany
FEATURES             Location/Qualifiers
     source          1..8018
                     /organism="Streptomyces fradiae"
                     /strain="T#2717"
                     /db_xref="taxon:1906"
     gene            1..936
                     /gene="urdL"
     CDS             1..936
                     /gene="urdL"
                     /note="UrdL"
                     /codon_start=1
                     /transl_table=11
                     /product="cyclase"
                     /protein_id="AAF00205.1"
                     /db_xref="GI:6002928"
                     /translation="MTTRSVEHEITVEAPAAAVYRLIAEVENWPRIFPPTIYVEHLER
                     GEGEERIRIWATANGKAKNWTSRRTLDADNLRTFRQEVSTPPVAAMGTWIIEPLSG
                     DSSRIRLLHDYRAVDDDPQGLKWIDEAVDRNSRSELAALKTNVELAHASEEITFSFED
                     TWWIDGSAKDYADFVNEAGLWVERLSHVASVRFSEDTPGLQSLEMDTLAKDGSTHTTK
                     SYRVTFPHIRIAYKQVTLPALMTLHTGVWTFTENESGTAASSQHTVVLNTENIAGILG
                     PEATVADAREFIRGALSTNSRATLGHAKDYAENKR"
     gene            939..2948
                     /gene="urdM"
     CDS             939..2948
                     /gene="urdM"
                     /note="UrdM"
                     /codon_start=1
                     /transl_table=11
                     /product="oxygenase-reductase"
                     /protein_id="AAF00206.1"
                     /db_xref="GI:6002929"
                     /translation="MVAPSLDVDVIVVGAGPVGLMLAGELLRTGGVRVTVLERLAEPT
                     TESRASMTMLTPSVEMELLHERGLVERLGPPPDAGPGHFGGIPLDLTEAGESRYAGQW
                     KAPQIRVEAVLSTWATELGAEVRRGHTVIGLVEAPDGVSVVATAPSGERLRLSAAYIV
                     GCDGEDSAVRRLAGFAFPGCADPTKELLRADLAGIELRERRFERHPNGVANARRGPDGI
                     GRVLLAGDAAHVQLPVGGQALNLGLGQDAMDLGGKLAAHITGKAGELLDTTEPATRWR
                     PAYSATSKHKPSCCSADPMWSLRAVFGELLGLGAARRHLASMISGLDGGAPTSVPRTG
                     PDATAHPGFTRQHTPHRRTTMGKLTGKTALVTGSSRGIGRATAIRLAREGALVAVHCS
                     RNREAADETVATIEKEGGRAFSVLAELGVPGDVHELFLALERGLKERTDATTLDILVN
                     NAGVMGGVAPEEVTPELFDRLVAVNAKAPFFIVQRAVTLIPDGGRIINISSGLTRFAN
                     PQEVAYAMTKGAMDQLTLHFAKHLGSRNITVNSVGPGITNNGTPVFDNPEAVAQMAGY
                     SVFNRVGEVTDVADVVAFLAGDDARWITGSYLDASGAPCSAESRG"
     gene            2945..4198
                     /gene="urdJ2"
     CDS             2945..4198
                     /gene="urdJ2"
                     /note="UrdJ2"
                     /codon_start=1
                     /transl_table=11
                     /product="transporter"
                     /protein_id="AAF00207.1"
                     /db_xref="GI:6002930"
                     /translation="MTTLSPRFDPHGRASAGPVARPRALLFVLAANMLIDALEVSVVL
                     VALPAIGGDLGLSPQDAQWMSGFALGFAALLLSGPRITARWGRRAYLVALLVFAVA
                     SSAGGLVHSGELLVLTRVIKGMCAALTAPLVGFCLIPDDRGPAPTASPGTGVTRF
                     GFTVGLLASGALTVLSWRWDLVLPAPIALGLMVLGFCLIPDDRGPAPTASPGTGVTRF
                     LRHGPIVRSALCAASLNGAYLGLLLLVTYQLHTGPGWNSWQTAVALLPACVPLMVSLP
                     FAGRMVGRLGAARLIVSGTLAATLGCAGCAVWGVSGVTAVLPAVALLVEGAGFVLSFA
                     ALNMQAVAGIAPESRQTAVSLYQTAVQLGAALTLPAVALLGSGGEGPYRTALLLITA
                     VAAVGAAVACTGTRRKAEQPCVSRR"
     gene            4195..4803
                     /gene="urdZ1"
     CDS             4195..4803
                     /gene="urdZ1"
                     /note="UrdZ1"
                     /codon_start=1
                     /transl_table=11
                     /product="epimerase"
                     /protein_id="AAF00208.1"
                     /db_xref="GI:6002931"
                     /translation="MSIPGAFLITPDQLSDERGAFYELRCDMLERAVGVPFQPQQINY
                     SVSKRHTLRGIHSVSIPPGQAKLVTCVRGALRDIVVDLRIGSPAFGAHQVTELDAVSG
                     RSVYVPEVWGTDSSRSPTTPVSATSSPSTYVPGTQIDINPLDPDLDLPWDCPQEPLIS
                     EKDAKASSLAEALASGTLPDLHDCRTSDAARTALVRAPSERQ"
     gene            4846..5943
                     /gene="urdGT2"
     CDS             4846..5943
                     /gene="urdGT2"
                     /note="UrdGT2; uses GTG start codon"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /protein_id="AAF00209.1"
                     /db_xref="GI:6002932"
                     /translation="MFALAPLATAARNAGHQVVMAANQDMGPVVTGVGLPAVATTDLP
                     IRHFITTDREGRPEAIPSDPVAQARFTGRWFARMAASSLPRMLDFSRAWRPDLIVGGT
                     MSYVAPLLALHLGVPHARQTWDAVDADGIHPGADAELRPELSELGLERLPAPDLFIDI
                     CPPSLRPANAAPARMMRHVATSRQCPLEPWMYTRDTRQRVLVTSGGSRVAKESYDRNFD
                     FLRGLAKDLVRWDVELIVAAPDTVAEALRAEVPQARVGWTPLDVVAPTCDLLVHAGG
                     VSTLTGLSAGEPQLLIPKGSVLEAPARRVADYGAATALLPGEDSTEAIADSCQELHAK
                     DTYARRAQDLSREISGMPLPATVVTALEQLA"
     gene            5968..7035
                     /gene="urdG"
     CDS             5968..7035
                     /gene="urdG"
                     /note="UrdG"
                     /codon_start=1
                     /transl_table=11
                     /product="dNDP-glucose synthetase"
                     /protein_id="AAF00210.1"
                     /db_xref="GI:6002933"
                     /translation="MKALVLAGGSGTRLRPFSYSMPKQLIPIANTPVLVHVLRNLREL
                     GVTEAGVIVGNRGPEISAVLGDGAEFGMRVTYIPQDAPRGLAHTVAITRDFLCDDDFV
                     MYLGDNMLPDGVVDMRGGIHRPSAGGQVVYKVPDPRSFGVAELGPQGEVLRLVEKPR
                     DPRSDMALVGYFFTAAIHDAVAAIEPSARGELETDAIQWLVSSGADVRASQYDGYW
                     KDTGNVEDVLECNRHLLDGLGARMDGTVDAASVLTGEVVIEPGAKVVRSRIMGAAIIG
                     AGTVVQDSHVGPHASIGRRCTVTDSRVENSIALDEASVTGVRGLRNSLIGRSASVGGT
                     GQENERYRLVGDHTRVEVAA"
     gene            7032..8018
                     /gene="urdH"
     CDS             7032..8018
                     /gene="urdH"
                     /note="UrdH"
                     /codon_start=1
                     /transl_table=11
                     /product="dNDP-glucose 4,6-dehydratase"
                     /protein_id="AAF00211.1"
                     /db_xref="GI:6002934"
                     /translation="MNILVTGAAGFIGSHFVRSLLADTYSGWEGARVTALDKLTYAGN
                     RNNLPPSNPRLEFVRGDVCDRALLRELLPGHHAVVHFAAESHVDRSLEGAGEFFRTNV
                     LGTQTLLDAVLDSGVERVHVSTDEVYGSIEQGSWTEDWPLQPNSPYAASKACSDLVA
                     RAYCAPTEVDLSITRCSNNYGPHQHPEKVIPRFVTNLLEGRQVPLYGDGRNVREWLHV
                     EDHCRGIHLVLNKGQAGEIYNIGGGNEYTNLALTEKLLELTGAGMEIRRVPDRKAHD
                     LRYSIDESKIREKLGYAPRISFEQGLSDTVAWYRDNPDWWKSIEHGGDSAAA"
BASE COUNT     1266 a   2927 c   2638 g   1187 t
```

ORIGIN

```
Query Match      12.3%;   Score 845.4;  DB 1;  Length 8018;
Best Local Similarity  64.7%;  Pred. No. 5.4e-56;
Matches 1311;  Conservative  0;  Mismatches 696;  Indels  18;  Gaps  3;
```

```
Qy    2267  GGTGAAGGCCCTGGTCCTGGCCAGTGAACCGGCAGCAGACTGAGGCCGTTCACCCACAC  2326
             | | |||||||||| ||| ||  || |||||||||||||| |||||||| ||||| || |
Db    5967  GATGAAGGCACTTGTGCTGGCCGGGCGATCTGCACCCGCGGCCCTTCAGCTATTC      6026

Qy    2327  CGCCCGCCAAGCAGCTGCTCCCCATCGCCCAAACAGCCGCTGCCTTCTTCTACGCCTGGAGTC  2386
             |||||  ||||   | |||  ||||||||||||||||  |||||| || |||||||||||
Db    6027  GATGCCCAAACAACTCATCCCCATCCCCGGCGGCGCTCGTCTGGGCGCGTACGGCCGGGGAA  6086

Qy    2387  CCTCGGCCGGCGGGGTCTCCGGAGGCGGCCTGCTCGTCGGGGCGCGTACGGCCGGGAGAT  2446
             ||| |||  ||| |||  |||||| || || ||| || | ||| | | |||| |||||
Db    6087  CCTCCGGAGCTGGGAGTCACCGAGCCGGTGTTCATCGTCGGCGAACGGGGGCCCGGAGAT  6146

Qy    2447  CCGGCGAACTCACCGGCGACGGCACCGGCGTTCGGGTTACGCATCACCTCCACCAGCC  2506
             | ||| |||||| | | || | || ||| |||||| ||| ||| |||| ||||||
Db    6147  CAGCCGCGGCTCCGGGCGAGCGGAGTTCGGCATGCGGTCCGTCACCGTCACCACCCAGGA  6206

Qy    2507  CCGCCCGCCTCGCTCGCCGACGCGGGTGCGCACCGGTGGGCGATCTCTGGGCGACGACGA  2566
             ||| || | | ||| | |||| ||| | || | ||| || |||||| ||  |||||||
Db    6207  TGCGCCGGTGGTCTCGCCGCCATACCGTGGGCCATCACCCCGCGCCGGCGCCCCGACGACGA  6266

Qy    2567  CTTCCTGCTGTGTACCTGGGGGACAAACTACCTGCCCCAGGGCGGTCACCGACTTCGCCCGCCA  2626
             ||| | |  ||||||| ||  |||| || || ||| || ||||| || | | |||| |||
Db    6267  TTTCGTGATGTACCTGGGCGACAAACATGCTGCCGACAGCGACGGTGTCGACATGCGACGCGGAGG  6326

Qy    2627  ATCGGGCCGCGATCCGGCGGCCGGCGGTGCTCGTCACCCCGGTCGGCGACCCGTCGGC  2686
             ||| |||  | ||| || ||||| ||||| ||| ||||||| ||| ||||||||||
Db    6327  AATTCACCGGCCCCCATCGGCCGACCCCCAGCGGTGTCGTGTACAAGGTCCCGGACACCGCGCTC  6386

Qy    2687  CTTCGGGTCGCGGAGTCGACGGCGGAACGTGCGACGGCGGCGGGAACCCGA  2746
             ||| |||||||||||| |||| |||||| | | ||||| ||||||| |||
Db    6387  CTTTCGCAGCCCCAATTCTTCGGCCTGCGTACGCCGGGTCGTGGGGAAGCCCCG  6446

Qy    2747  CGTCCCGGCAGCTCGCTCGCCGTACGGCGTATGGGGCCCCGTCGGCCCCACGA  2806
             || ||||||||||||| |||||||||||||||||| | | ||| || ||||
Db    6447  CGACTACTGGCGACGCGACAACGTCGGCGGTGCAATCGAGCTCGTGGGGGC  6506

Qy    2807  GGCGGTACGGCCGCCATCACCCCCTCCGCCCGCGGCGGCGAGCTCGATGAGATCACCACGCCGTGCA  2866
             | || |||  || | || |||  | |||| |  ||| ||| ||| ||  |
Db    6507  CGCCGTCGGCGGATCCAGCCCAGTGGGCCGCGGGCGCGGCGAACTGGAGACTGGAGATCGAACC  6566

Qy    2867  GTGGATGATCGGCGCCGGGGGCGGTGCCGCGTACGGCGGGCGGAGACCACCACCGCCTGGCGCCGA  2926
             || ||| |||| |||  ||| ||  |||  |   ||| | || |||||
Db    6567  GTGGCTGGTCTCCGGCGACGTCGCGCGCCGACGGCCTCAGTACGGCCGGGCTCCGGGCGGAGGA  6626

Qy    2927  CACCGGCAGCTCGGGAGGACATGCTGGAGGTCAACGGTCTGAGGGACTGGAGGG  2986
             ||| |||||| || ||| | ||| ||||| ||| || |  | |||||| ||
Db    6627  CACCGGCCAACGTGCTCAACAAGGGCAGGCCGACTCACGACCCCCCGGTGGGGC  6686

Qy    2987  CCGCATCGAGGGGAAGTGCACGCGCACACTCGGTGCTCGGTCGGGTGGCCGA  3046
             |||| || |  |||| ||  | || |  | |||||| ||| |||| |||||
Db    6687  GCCGGCAACGAGTACCACCACCAGTGCTGCCCCAGTGTGCTGACCGGTGATCGAACC  6746

Qy    3047  AGGCGGCATCGTGCGGGGCTCACACGGTGTCATCGGCCCGGTGTGATCGGCCGCCGT  3106
             |||| ||| |  || ||| ||| ||||| ||||| || ||  ||||  ||||||||
Db    6747  CGGGGCGAAGGTCTCCTCCGGCGCCCCGCATCGGGGCGGCCGTACGCGGCCGGCCGT  6806

Qy    3107  CGTCAGCAACTCCAGTGTCGGCGGAGGACTGCTCCATCGGGGGCGTCCGGAGGA  3166
             | || ||| ||||  |||| ||||||||||| |  |||| ||||| ||| |||
Db    6807  CGTGCAGGACGACGCCACCTGGGTCCTGGTGCAATCGACGGGGGCTGGGGGGC  6866

Qy    3167  CAGCGCCATCGAGTACTCCGTTCCGTCCGTGCGCGCGGCCGTCCCCGCAT  3226
             ||| ||||||  ||||||| || ||||||  |  || | |||  ||| |
Db    6867  CAGCGCCGTGGAGAACTCCATAGCCCTCGACGAGGCGCGTCTGACGCTCCGAGGACT  6926

Qy    3227  CGAGGGCGTCCCCATCATCCGGCGCGTCTGCGGCCCCGTCTCCCCGCAGGC  3286
             ||||| ||  ||  || |||||||  ||||| || |||| || ||||| |
Db    6927  GCGCAACTCGCTCATCGGGCGCTCGGCGCGTCCGGCGCACCGGCCAGGAACGAGCG  6986
```

```
Qy    3287  TCACCGACTGGTGATCGGGCGCCAGCACCAGCAGGTGTATCTCACCCCATGACCACGACCATC  3346
             |||||| || |||  ||||  |||||||||||||  | ||| |||||| | |  | ||||
Db    6987  TTACCGCCTGGTCGTCGGAGACCACCACCACCCGAGTGGAGGTCGC------AGCATGAACATC  7040

Qy    3347  CTCGTCACCGGCGGAGCGGCGGCTTCATTCGGTCGCCGTACGTCCGCGGCTCCTGGCCC  3406
             || ||||| || | ||| |||| || || |||||||||||||| |||| |||||||||
Db    7041  CTCGTCACCGGGCGGCGGCGGCCGGTTCATCGGCTCCGCCGCACTTCCGTCCGCGGCCGAC  7100

Qy    3407  GGGGCCCCGGC------GGCGTGCGGTGACCGTGCTGCTCCTGCACAAACTCACCTACGCCCGGC  3460
             ||| ||  ||        ||||| | |||| || || |||||| | | ||| |||
Db    7101  ACCTACTCCGGCTGGGAGGAGGAGCGCGGTCACTGACCGTCTGGGCTCACAAACTGACCTACCGGGC  7160

Qy    3461  AGCCTTGCGCCCGTCGTGCGACGCGGTGCTGGGCGCGTACGGCCGGGGAGAT  3520
             | || | || || |  || ||||||| |||||| | | || || |||||
Db    7161  AACCGGAACAAACTGC------CGCGCGCGGGCCTGGCCGCCGGCGTTCGTCGGGGGCGGAC  7214

Qy    3521  GTGTGCGACACCGGCGTCGTCGACAGCTGCGACACGCTGCTGGCCCACCGACCAGCC  3580
             | ||||| || | |  || | | | ||| || |||||| |  | || ||| | |
Db    7215  GTGTGCGACGCCGCCCTGCTGCGGAGCTGCTGCCGCGGCGCCACCACGCCGTGGTCCATTTC  7274

Qy    3581  GCGGCCGAGTCGCACGTGCCGTCCATCACCGACAGCCGGTGCCTTCACCCGCACCAAC  3640
             | |  || |  ||||||||   ||  |||| || |||| || ||| || ||||||||
Db    7275  GCGGCGGAATCCCACGTGGACCGCGTCGCCGAGTTCTTCCGCACGCGGGAGTCGAGGAAC  7334

Qy    3641  GTGCTGGGCACCCAGGTCGTCGCTCGTGCTGCGACGCTCCCGCCGCAGCTTCGGTG  3700
             | || |||||||||||| |||||| | ||| | ||| |||||| || | |||| |
Db    7335  GTCCTCGGTACGCGACACTCCTCGACCGCTGGCAGCGGCGAGTGAGCGGGTCGTCGAGTCGTCGTT  7394

Qy    3701  CACGTCTCACCGACGAGGTGTACGGCGGGCCGCGGCGGAGAGCGAC  3760
             |||||||| ||| | |||| |||| ||||| |  |||| |||||
Db    7395  CACGTCTCCACCGACGAGGTGTACGGCTCTCGATCGAGAGACCGAGGACTGG  7454

Qy    3761  CCCCTGCTTCCGACCTCGCCGCCGGGGCTACGGCGCTAGCAGGCCCTCATGCCGCGCTC  3820
             || || | || ||| | |  ||||| | | |||| | |||||||||||| |||||
Db    7455  CCGCTGCAGCCCAATTCTCCGGCCTGCGTACGCCGGGTCGTGGTCGGTCGCGCGA  7514

Qy    3821  GCCCACCACCACCCACCGTTGTCGACGTGTGACCGCGTGTTCGAACAACTTCGGC  3880
             || | ||| ||||| ||| | | |||| ||||| ||| | | |||||| ||||
Db    7515  GCCTACTGGCGACGACGTGCTCAACAAGGGTCGACCTCTCCATCAACAACTACGGC  7574

Qy    3881  CCCCACCAGGATCCCCGGCGGCGGCGGAGATCCATACCGGCGCGCTTCCTCGTCCGCGGGCGGC  3940
             || || |||| | || ||| ||||| | ||  |  | |||| || |||  || || |||| |
Db    7575  CCGACCAGCAGCCCTGAGAAGGTCATTCCCACGACGTCGTCACGGAACCTCTTCGAAGGACGC  7634

Qy    3941  ACCGTTCCCCTCTACGGCGACGGGCGCGTGCCACGTGCCGACGACCAC  4000
             ||| | || ||| |||| |||||  |   || || || | |||| ||
Db    7635  CAGGTCCCCGTGTACGGCTACGGCGACGGCCAATGTGCCGCCGGGGACCAC  7694

Qy    4001  GTCAGGGCCGTCGAACTGCGGGCCGTGGGTGCGGGAGAGATCTACAACATCGGG  4060
             || | ||| |  |||||| |  ||  ||||| | |||||| | ||||| ||||
Db    7695  TGCCGCGGTATCCACCTGTGCTCAACAAGGGCAGGCCGGGATCACAACATCGGG  7754

Qy    4061  GGCGGCACCTGCTTGCCCAACTGGAGCTTCACGCACCGGTTGCTCCGGCGCG  4120
             |  || |||||| |||| ||  ||||  |||  |||| ||| |||  || | |
Db    7755  GGGGGCAACGCAACGAGTACCACCAACCTCACCGAGAAACTGCTGCCGGGTGCC  7814

Qy    4121  GCGCCCGGAGCGCCATCGTCACGTCCACGGGCGGCTCAGGACGACCGCTACCGGTC  4180
             ||| ||| |  |  |  |||||| ||||   |  || || || || | |||| ||
Db    7815  GGCCCGGAGATGATCCCCCGCGTCCCGCGACAGGCCAAGCTGCACGCCGTACTGGTC  7874

Qy    4181  GACCACAGCAAGAGATCACCGGCACCCGCACCGGTTACCGGACCGCGG  4240
             ||| | | ||| | |  ||| | ||  |||||||||  |||| | |||
Db    7875  GACGAGTCCAAGATCCAAGAGATCGGCTGCCCCCGCGGATCAGCGTTCGAACAGGGGC  7934

Qy    4241  CTGGCGACACCCGGAGTGTACGAGCGGCGGCGACGGACTCGTGG  4285
             ||| ||||||||||| ||||||
Db    7935  CTGTCCGACACCGTCGCCTCGTGCTTGGTACCGCCGACAACCCGGACTCGGTGG  7979
```

RESULT 15
STMSVIDEOX/c
LOCUS     STMSVIDEOX         7365 bp    DNA      linear     BCT 16-MAY-1996

```
DEFINITION  Streptomyces violaceoruber Tu22 dTDP-glucose dehydratase (graE)
            gene, complete cds, dTDP-glucose synthetase (graD) gene, complete
            cds, ORF4 (Streptomyces coelicolor mmr homolog), complete cds.
ACCESSION   L37334
VERSION     L37334.1  GI:763509
KEYWORDS    dTDP-D-glucose 4,6-dehydratase; dTDP-glucose synthase; transport
            protein.
SOURCE      Streptomyces violaceoruber (clone pOJ446-22-24, strain Tu22) DNA.
ORGANISM    Streptomyces violaceoruber
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 7365)
AUTHORS     Bechthold,A., Sohng,J.K., Smith,T.M., Chu,X. and Floss,H.G.
TITLE       Identification of Streptomyces violaceoruber Tu22 genes involved in
            the biosynthesis of granaticin
JOURNAL     Mol. Gen. Genet. 248 (5), 610-620 (1995)
MEDLINE     96027933
PUBMED      7476861
FEATURES             Location/Qualifiers
     source          1..7365
                     /organism="Streptomyces violaceoruber"
                     /strain="Tu22"
                     /db_xref="taxon:1935"
                     /clone="pOJ446-22-24"
     CDS             complement(151..816)
                     /note="ORF1; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA99938.1"
                     /db_xref="GI:763510"
                     /translation="MLCDSGIGLSPFVILHLHLHHGTEFTMPALRSRTRTAAIAVPLA
                     LAGLMATTSGAQAAPVKTIYFDCRVGTVTGLAYSIYVDGVAPAHYRPGATFDVILDPAP
                     ITPNPAYNKEVKNVALQFKLPAGAKVLSYEVADRLAAGTPELQIDGTKAVLRAAGPYV
                     ANAPFDLPKITIKLKAPTTRGTLETPFGGTSHDDPGFTWTFISPNWTARPSSPAGPTS
                     RSR"
     gene            complement(889..1854)
                     /gene="graE"
     CDS             complement(889..1854)
                     /gene="graE"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="dTDP-glucose dehydratase"
                     /protein_id="AAA99939.1"
                     /db_xref="GI:763511"
                     /translation="MRLLVTGAAGFIGSHYVREILAGSYPESDDVHVTVVDRLIYAGR
                     RDNLPEHHERLDFYHGDICDRDLLDRVLPGHDAVVHFAAESHVDRSLTGPEFVRTNV
                     MGTQQLLDAALHAGVDRLVHVSTDEVYGSLDSGTWTEDSPLLPNSPYAASKASTTWSA
                     APTTVRHGLDVRITRCSNNYGPRQHPEKLIPNFVTRLLLTGRQVPLYGDGRNVREWLHV
                     DDHCRALQLVLTKGRAGEIYNIGGGSGMSNREMTARLLLDLLGADWDMVRHVEDRLGHD
                     FRYAIDDSKIREELGYAPRWSIESGLGAVVDWYRDHPDFWRAPAS"
     gene            complement(1851..2909)
                     /gene="graD"
     CDS             complement(1851..2909)
                     /gene="graD"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-glucose synthetase"
                     /protein_id="AAA99940.1"
                     /db_xref="GI:763512"
                     /translation="MKALICPEDGDRLRPFTYSMPKQLVPVANKPILVHCLENVRAIG
                     VEEVAVVGGRDADEVRAVVGDGSAFGLDVTYLQQEAPLGLAHCVSIAEEFLGDEDFVM
                     YLGDNILAEGIAEASARAFRDERSAARLLLTKVADPRAYGVAETDATGRVHALVEKPER
                     PRSDLAVIGVYFFAVRDAVRAIEPSARGELEITDAIQYIVERGDRVVADEYTGYWK
                     DTGSPDDLLDCNRVLLGRLRPGVHGEVDAASTVEGTVVVEAGRSWSDSRLVGPLVVGA
                     GSVVRGSELGPYTALGRDCVLEDAGIRDSIVLDGVSIQGVRGLSGSLIGRSAAVRTGE
                     AAGRLLIGDHTQAEVAA"
     RBS             complement(1862..1866)
                     /gene="graD"
     RBS             complement(2922..2925)
     CDS             complement(3279..4952)
                     /note="ORF4; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA99941.1"
                     /db_xref="GI:763513"
                     /translation="MTDTANRRSATPDSGADAVDKRRIRLIMTGLLLGLFMAALDQT
                     IISAALRTIADDLNGLSEQAWANPSYMITSVIMTALYGKLSDIYGRRPVYCTAVGVFV
                     LGSVLCGLAQSMTILAVFRGQGIGAGGLMSLAFAILTDLVPLAERSRYQAWFGAVFGV
                     SAVVGPVAGGFFAGLDSFLGASGWRWAFFINVPIGVAGALIIAALVRVQPGRAQHKFD
                     IAGLLALIACLLPLLFAVEQGPAWGWTDPGTLVLFGLGAVGLVLFSLAQKRAADAALL
                     PTPMFRNPLFSVYNGVNVLVGAAVFGALSVLPLIQMVKGLSPTQAGLMMLPQTIGIV
                     VAGRIARPYVTRTGRYKAVLLSGLVLMVVATFWFGVLTADTSLMQTGVAAGVMGFGIG
                     LCWQVMLIAIQTGVAPQYMGAGMGSFTFRQIGGTGRHRVFLSMFFGAVGEKVAAAYR
                     GAASDPAYTAAVNDPAVTGQAANKVLLGFGRRVGLDNSSFLDHADPRLAKPFLEGLAE
                     AMQTVFVVSCVMLLIALVLAAVPFKEKPPRAGRRFAGRRKETGKQPAAKK"
     RBS             5413..5416
     CDS             5423..6370
                     /note="ORF5; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA99942.1"
                     /db_xref="GI:763514"
                     /translation="MPLVAATGLPGAPITSRTMEDFMQRDRHGNPLEIPKDPHERNLF
                     NGRGMARLALGSMEGLVPHVERWQPDVLVAGLSYAPLVAHRFGLPWVRHALNMGEP
                     SIIDLSAAAELAPELEEMGLSAIPEPDMVEICPPGARRPDAGPAQFMRVVPFNTQRA
                     LEPWMYAKGDRPRVLVSAGSRVTADYEADALSALVEKVAGLDVELLIAARPQEIADALG
                     DLPDNVRAGCVPLDVVLRTCDLLVHRAGNMHAIVCGVPQLVIPAMPKQVGMSARL
                     AEYGAAIMLTAGQDDSPENVPRPAGSCWRTRRTRPGPTS"

BASE COUNT     1098 a   2580 c   2668 g   1019 t
ORIGIN

Query Match             12.8%;  Score 834.6;  DB 1;  Length 7365;
Best Local Similarity   65.0%;  Pred. No. 3.7e-55;
Matches 1306;  Conservative    0;  Mismatches  679;  Indels   24;  Gaps    4;

Qy   2298 GGCAGCAGACTGAGGCCGTTCACCCACCACGCCGCCAAGCAGCTGCTCCCCATCGCCAAC 2357
               || || ||||| ||  ||||| ||| ||  | || |||| || ||||||||||
Db   2882 GGGGACAGGTTGCGTCCCTTCACCTATTCCATGCCCAAACAACTGGTCGCCGTCGCCAAC 2823

Qy   2358 AAGCCCGTGCTCTTCTACGGCCACTGCCTGGAGTCCCTCCCGGCGGGTGTTCCGGAGGCCGGC 2417
               |||| ||||| | ||||| | | |||| |||| ||| ||||  ||||| |
Db   2822 AAGCCGGATCCTCGTGCCACTGCCTGGAGAACGTCCGTGCCATCGGTGTCGGAGGAGTCGCC 2763

Qy   2418 GTCGTCTGGGGCGGCGGTACGGCCGGGAGATCGCCGAACTCACCGGCGACGGCCACGCGTTC 2477
               |||||||||| | || || |  | | ||| || || ||||| ||| | | ||||||  |||
Db   2762 GTCGTCGGCGGACCCGCGACGAGGTCCGGCGTGGTCGGTGACGGCGTCCGGCGTTC 2703

Qy   2478 GGGTTACGCATCACCTACCTCCACAGCGCCGCCAATCGGCCGGCCGATCCGGCGCTG 2537
               |||| | |||| ||||| || | ||  |||||| || ||| || |||| | ||||
Db   2702 GGCCTCGACGTCACCTACCTCCAGGAGGCCCGTCGGCCCTCCACTGCGTCTCC 2643

Qy   2538 ATCGCCCGGCGGCTTCTGGGCGACGACGACTTCCTGCTGTACCTGGGGGACAACTACCTG 2597
               |||| |||||||||| ||| ||||| ||||| ||| |||| || ||||| |||
Db   2642 ATCGCCCGAGGAGTTCCTCGTGACGAGGACTTCGTGATGTACCTGGGGCGACAACATCCTC 2583

Qy   2598 CCCCAGGGCCGTCACCGACTTCGCCGCCAATCGGCCGGCCGATCCGGCGCTG 2657
               | |||||| ||| ||  ||||| |||| |||| || ||||| | ||||||
Db   2582 GCTGAGGGCATCGCCGACGAACGTTCGCGCCCTCCGGACGGG 2523

Qy   2658 CTGCTCACCCCGGTCGCGGACCCGTCGCCGCCTTCGGCGTTCGAGGTCGAGCGGACGGG 2717
               |||| |||| |||| |||||| || |||| || ||||| | ||| |||  || |
Db   2522 CTGCTCACCAAGGTGCCGCCGTGTCCTACGGCGTCGCCGAGACGGATGCCACCGGC 2463

Qy   2718 AACGTGCGGCTTGGAGGAGAAACCCGACGTCCCCGGCGCTCCGCCGCTCGCCATCGGC 2777
               ||||| ||||| || || | || ||| |||| |||||| ||| ||| |||
Db   2462 AGGGTCCACGCGCTGGTCGAGAAGCCCGAGAGGCCGCGCAGCGACCTTGCTGATCGGC 2403

Qy   2778 GTGTACGGCCTTCAGCCCGGCGTTCACCGAGCGGTACGGCGCCATCACCCCCTCCGCCCGC 2837
               ||||| |  || ||| |||| |||| ||||| || |||| |  |  | |||| ||||  |
Db   2402 GTGTACTTCTTCGCCGTCCGCGATCGAGCCGAGCGGGAACTGGAGATCACGGACGCCATCCAG 2343

Qy   2838 GGCGAGTCGAGATCACCCACGCCGTCAGTGGATGATCGCAGTGGTTGTACGG 2897
```

```
Db  3972 TGCCGACTGGCTGCACGTCGACGACGACCACGTCAGGGCCGTCGAACTCGTCCCGCGTGTCGG 4031
         |||||| ||||||||| || |||||||| |||||| | ||| ||||| |||||
Qy  1220 TGCCGAGTGGCTGCACGTGGACGACCACTGCCGGGCGCGTGCAACTGGTGCTCACCAAGG 1161

Qy  4032 GCCGGCCGGGAGAGATCTACACATCGGGGGCGGCACCTCGCTGCCAACTGGGAGCTCA 4091
         |||| ||| |||||||||||||| || |||||||| || |||||| ||||
Db  1160 GCCGGCCGGGCGAGATCTACAACATCGGGGGTGGCAGCCGGTATGAGCAACCGCGAGATGA 1101

Qy  4092 CGCACCGGTTGCTCGCACTGTGCGGCCGCATCGTCGGAGGCGCATCGTCCACGTCGAGAAC 4151
         | |||| ||||| | || |||| ||| |||| || |||||||| | |||||||
Db  1100 CCGCCCGCCTGCTGCGACCTCCTGGGGGCCGGACGTGGGCCCATGGTGCCGTCGGAGGACC 1041

Qy  4152 GCAAGGGGCACGACCGGCCGCTACGCGGTCGACTCCAGTCGGTCCGGCCCCGTACACC 4211
         || || |||||||| ||||| ||||| |||||||| || ||||| ||||
Db  1040 GCCTCGGGCACGACTTCCCGCTATGCGATCGACGACTCCAAGATCCGCGAGGAGCTGGGGCT  981

Qy  4212 ACCGGCCGCGCCACCGACTTCGCGACCGCTGGCCGCCGACACCGCGAAGTGGTACGAGCGGC 4271
         || ||||||||| |||| |||||| || |||| ||||||
Db   980 ACGCGCCCGCCTGGTCGATCGAGAGCGGGTCTGGGGGCCGGGTCTACCGCCGACC  921

Qy  4272 ACGAGGACTGGTGCGTCCCCTGCTCGCC 4300
         || ||| ||||||| ||||| ||||
Db   920 ACCCCGACTTCTGGCGCGCCCGGCCTCC  892

Search completed: May 28, 2003, 19:47:19
Job time : 12493 secs
```

```
Db  2342 GGCGAACTGGAGATCACCGACGCGATCCAGTACCTCTGTGGACGGCGCGGCACCGGGTCGTC 2283
Qy  2898 GCCGAGACCACCACCCGCCCTGGGCGCGCAGCGCGGAGGACATGCTGGAGGTC 2957
Db  2282 GCCGACGAGTACCGGCTACTGGAAGGACACCGCAGCCCGACGACCTGCTGGACTGC 2223

Qy  2958 AACCGTCACGTCCTGGACGCTGGAGGCGATCGAGGGGAAGGTGACGGCCGCACAGC 3017
Db  2222 AATCGGTACTCCTCCGGCCGGCGCTCCCGGCGTACACGGCGCGGAGGTCGACGGCCAGT 2163

Qy  3018 ACGCTGGTCGGGCGGCGGTCCGGGTGGCCGAAGGCGGCGATCGGTGCGGGGGTCACACGTGGTG 3077
Db  2162 ACCGTCGAGGGGACCCGTCGTCGAGGCGGACGGTCGTGGAGCGACTCCGCGCCTGCGTC 2103

Qy  3078 GGCCCGGTGGTGATCGGCGGCGGGTGCCGTCGTCAGCAACTCCAGTCGTCGGCCCGTACACC 3137
Db  2102 GGACCGCTCGTCGCGGCCGCAGCGTCGTCGCGGCGAGCGTCGGCGCCCGTACACC 2043

Qy  3138 TCCATCGGGGAGGACTGCCGGTCGAGGACAGCGCCATCGAGTACTCCGTCTGCTGCGC 3197
Db  2042 GCTCTCGGCCGCGACTGCCGTGCTGGAGGACGGGGGATCCGCGACTCCATCGTGCTCCAAC 1983

Qy  3198 GGCGCCCAGTCGAGGGGCGTCCCGCATCGAGGCCGTCCTCATCGGCCGCGGCGCGCCGTC 3257
Db  1982 GGGGTGTCGATCCAGGGCGTCGCGCTGCGCTCGCTCATCGGCGTCGGCGGCCGCC 1923

Qy  3258 GTCGGGCCCGGCCCCCCGCTCTCCCGCAGGCTCACCGACTGGTGATCGGCGACCAGCAAG 3317
Db  1922 GT------GCCCACCGGCGAGGCGGCCGGCGACGGCTCATCATCGGGGACCATACCCAG 1869

Qy  3318 GTGTATCTCACCCCATGACCACGACCATCCTCGTCACCGGCGGAGCGGCTTCATTCGCT 3377
Db  1868 GCGGAGGTGGCGGCAT------GAGGCTCCTGTCACCGGCGGCGGCCCGGTTCATCGGCT 1815

Qy  3378 CCGCCTACGTCCGCCGCTGTCG------CCCGGCGGGCCCCGGCGCGGTCGCGGTGA 3431
Db  1814 CGGCACTACGTGCGCGAGATACTGGGCGGGAGCTACCCGGAGTCCGACACGTCCACGTCA 1755

Qy  3432 CCGTCCTCGACAAACTCACCTCGCGGCGGCCGCAGCGGCCTCGCCGCCTCGACACGGTCGTGACC 3491
Db  1754 CGGTCGACACCGGCCTCACCTACGCGGCGGCGCGGACGCTGCCCGAG------CACC 1701

Qy  3492 ATCCCCGGCCTCACCTTCGTCAGGGCGACGTGTGCGACGTGCGACACGCCGCTCGTCGACACGTGG 3551
Db  1700 ACGAACGGCTGCGACTTCGTCCACGGGCAGCATCTGCCACGGGCGACACCTCCTGGACCGGGTCC 1641

Qy  3552 CCGCGCGGCACGACGACATCGTGCACTTCGGCGCGGCCGAGTCGACCGTCGCACGTCGACCGTCCATCA 3611
Db  1640 TCCCCGGCCACGACGCGGTCGTGCACTTCGCGCCGCCGAGTCGCCGCTCCCCTCA 1581

Qy  3612 CCGACAGCCGGTGCCTTCGACCTCATGGCCTCGCCGCCACCAACGTGCTGGCACCGCCAGTCCTCGACGCCG 3671
Db  1580 CCGGCCCCGCCGCAGTTGCTCGCCGTACGGGCGCGGCCGGCAGCTGCCGACCTCGACGCGCG 1521

Qy  3672 CGCTCGCCACCAGCGTGTGCGCGACGTGCGCCACCTTCGTCGCGACGCTCTCCACCGACGAGTGTACGGCTCCC 3731
Db  1520 CACTCCAGCGCGTGGACCGAGTGGACCGAGAGTGTACGGCTCGT 1461

Qy  3732 TCCCCGCACGGGCGCCGCCGGAGAGCACCCCGCTTCGGACGCCGACCTCGGCCGGCGCGCGT 3791
Db  1460 TGGACACGGCGGCACCTGGACCGGCGTCGTTCGCACTTCGCCCGAACTCCCCCTACGCGCGCGT 1401

Qy  3792 CGAAGGCCGGCCGCTCGGACCTCATGGGGCTGCGCGGACGCCGCACCCACGGCCTTGGACGTCC 3851
Db  1400 CCAAGGCCTCCACGACCTGGTCGCGCCGCCTACCACCGCGACCGCCACGGTACGCCGTCC 1341

Qy  3852 GGGTGACCGGCTGTTCGAACAACTTCGGCCCCCCCACCAGCATCCCGACGAGCTCATACCGGC 3911
Db  1340 GCATCACCCCGCTGCTCCAACAACTACGGCCCCGCCAGCACCCCGCCGCGAGAAGCTCATCCCGA 1281

Qy  3912 GCTTCCTGCACCAGCCTCCCTGTCGCCGGCACCGTCCCTCACGGCCGAGGGCGGCGCACG 3971
Db  1280 ACTTCGTCACCCGGTCGCTGCACGACGGCGGCAGGTTCCGGTGCTGCGACCGACGCGCAACG 1221
```

```
               GenCore version 5.1.4 p5_4578
          Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   May 28, 2003, 15:28:00 ; Search time 858 Seconds
                                    (without alignments)
                             17989.739 Million cell updates/sec

Title:         US-09-922-683-7
Perfect score: 6854
Sequence:      1 CTGCAGGGTTCCTGGTGCA.........TCGAGCCGATCCGGCTGCAG 6854

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters:          4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     N_Geneseq_101002:*
        1:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
        2:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
        3:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
        4:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
        5:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
        6:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
        7:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
        8:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
        9:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
       10:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
       11:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
       12:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
       13:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
       14:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
       15:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
       16:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
       17:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
       18:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
       19:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
       20:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
       21:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
       22:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
       23:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
       24:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| | Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| | 1 | 6854 | 100.0 | 6854 | 19 | AAT76903 | S. glaucescens Pst |
| c | 2 | 877.8 | 12.8 | 109519 | 22 | AAS08693 | Micromonospora DNA |
| c | 3 | 817.2 | 11.9 | 14186 | 23 | ABL50559 | Micromonospora car |
| c | 4 | 745 | 10.9 | 38064 | 24 | ABA99469 | Actinoplanes sp SE |
| c | 5 | 739.6 | 10.8 | 1350 | 24 | ABA99445 | Actinoplanes sp DN |
| c | 6 | 531.6 | 7.8 | 3291 | 21 | AAZ56004 | Contig 003 from co |
| | 7 | 529 | 7.7 | 3292 | 21 | AAA75636 | Nucleotide sequenc |
| c | 8 | 527.4 | 7.7 | 12441 | 21 | AAZ87284 | S. venezuelae deso |
| c | 9 | 527.4 | 7.7 | 13613 | 21 | AAZ87319 | S. venezuelae deso |
| c | 10 | 527.4 | 7.7 | 13613 | 24 | AAD39043 | Streptomyces venez |
| | 11 | 522.6 | 7.6 | 546 | 19 | AAT76906 | S. glaucescens acb |
| | 12 | 497.6 | 7.3 | 2634 | 14 | AAQ39093 | Streptomyces nodos |
| c | 13 | 486.8 | 7.1 | 16020 | 21 | AAA39283 | Streptomyces nogal |
| | 14 | 465.8 | 6.8 | 990 | 22 | AAF88341 | S. spinosa DNA fra |
| | 15 | 465.8 | 6.8 | 2310 | 20 | AA21502 | Saccharopolyspora |
| | 16 | 465.8 | 6.8 | 2363 | 22 | AAF88340 | S. spinosa polyket |
| | 17 | 454.4 | 6.6 | 3035 | 21 | AAA63462 | Streptomyces globi |
| | 18 | 454.4 | 6.6 | 42000 | 21 | AAA63349 | Streptomyces globi |
| | 19 | 454.4 | 6.6 | 63164 | 21 | AAA63348 | Streptomyces globi |
| c | 20 | 454.2 | 6.6 | 1476 | 24 | AAS18443 | Contig 122 DNA enc |
| | 21 | 447 | 6.5 | 1014 | 21 | AAZ87289 | S. venezuelae deso |
| | 22 | 447 | 6.5 | 1014 | 24 | AAD39047 | Streptomyces venez |
| c | 23 | 446.8 | 6.5 | 14806 | 22 | AAD03809 | Streptomyces galil |
| c | 24 | 433 | 6.3 | 9994 | 22 | AAC85191 | S. avermitilis 10 |
| c | 25 | 432.4 | 6.3 | 1074 | 24 | ABA99444 | Actinoplanes sp DN |
| | 26 | 432.4 | 6.3 | 38064 | 24 | ABA99469 | Actinoplanes sp SE |
| c | 27 | 387.6 | 5.7 | 2219 | 17 | AAT36646 | Acarbose biosynthe |
| c | 28 | 387.6 | 5.7 | 17955 | 19 | AAV56642 | Actinoplanes sp. a |
| | 29 | 382.2 | 5.6 | 1026 | 24 | ABA99459 | Actinoplanes sp DN |
| c | 30 | 276.2 | 4.0 | 541 | 19 | AAT76907 | S. glaucescens Hst |
| | 31 | 273 | 4.0 | 349980 | 22 | AAH41225 | Pyrococcus abyssl |
| | 32 | 272.2 | 4.0 | 35133 | 23 | ABL50991 | Thermus caldophilu |
| | 33 | 272.2 | 4.0 | 35134 | 23 | ABL50990 | Thermus caldophilu |
| c | 34 | 267.6 | 3.9 | 21719 | 23 | AAS59568 | Propionibacterium |
| | 35 | 248.4 | 3.6 | 114955 | 20 | AAX53491 | Human adenosine Al |
| | 36 | 238 | 3.5 | 4403765 | 22 | AAI99683 | Mycobacterium tube |
| | 37 | 236.2 | 3.4 | 28804 | 17 | AAT37329 | Sphingan biosynthe |
| | 38 | 235.4 | 3.4 | 114955 | 20 | AAX53491 | Human adenosine Al |
| | 39 | 234.8 | 3.4 | 28804 | 18 | AAT92474 | Sphingomonas genus |
| | 40 | 234.8 | 3.4 | 28804 | 20 | AAV99812 | Sphingomonas S88 s |
| | 41 | 234.8 | 3.4 | 28804 | 20 | AAV81474 | Chromosomal fragme |
| | 42 | 232.2 | 3.4 | 1200 | 21 | ABK73322 | cDNA encoding UDP |
| c | 43 | 224 | 3.3 | 4776 | 20 | AA23707 | S. erythraea-pFL37 |
| | 44 | 205.2 | 3.0 | 1131 | 22 | AAH65338 | C glutamicum codin |
| | 45 | 205.2 | 3.0 | 1140 | 22 | AAF71585 | Corynebacterium gl |

ALIGNMENTS

```
RESULT 1
AAT76903
ID   AAT76903 standard; DNA; 6854 BP.
XX
AC   AAT76903;
XX
DT   02-JUL-1998  (first entry)
XX
DE   S. glaucescens PstI DNA fragment containing acarbose biosynthesis genes.
XX
KW   Acarbose biosynthesis; acbA gene; acbB gene; acbC gene; acbD gene;
KW   acbE gene; acbF gene; enzyme; alpha-amylase inhibitor; treatment;
KW   diabetes; ds.
XX
OS   Streptomyces glaucescens.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..714)
FT                   /*tag= a
FT                   /product= acbA
FT                   /note= "acarbose biosynthesis gene as given in the
FT                   specification"
FT   CDS             complement (715..2006)
FT                   /*tag= b
FT                   /product= acbB
FT                   /note= "acarbose biosynthesis gene"
FT   CDS             2258..3331
FT                   /*tag= c
FT                   /product= acbC
FT                   /note= "acarbose biosynthesis gene as given in the
FT                   specification"
FT   CDS             3332..4306
```

FT          /*tag= d
FT          /product= acbD
FT          /note= "acarbose biosynthesis gene"
FT   CDS    complement (4380..5414)
FT          /*tag= e
FT          /product= acbE
FT          /note= "acarbose biosynthesis gene"
FT   CDS    5676..6854
FT          /*tag= f
FT          /product= acbF
FT          /note= "acarbose biosynthesis gene"
XX
PN   DE19622783-A1.
XX
PD   11-DEC-1997.
XX
PF   07-JUN-1996;   96DE-1022783.
XX
PR   07-JUN-1996;   96DE-1022783.
XX
PA   (FARH ) HOECHST AG.
XX
PI   Decker H;
XX
DR   WPI; 1998-033827/04.
DR   P-PSDB; AAW23831; AAW40381; AAW40382.
XX
PT   Recombinant DNA molecule comprising genes for biosynthesis of
PT   acarbose - an alpha-amylase inhibitor useful in treatment of
PT   diabetes
XX
PS   Claim 4; Page 14-28; 35pp; German.
XX
CC   This DNA sequence is a 6.8 kB PstI fragment isolated from Streptomyces
CC   glaucescens GLA.O and encodes the genes, acbA, acbB, acbC, acbD, acbE
CC   and acbF which are involved in the acarbose biosynthesis pathway. These
CC   genes and encoded enzymes are useful for producing acarbose, which is an
CC   alpha -amylase inhibitor useful in the treatment of diabetes.
XX
SQ   Sequence 6854 BP; 1018 A; 2445 C; 2445 G; 946 T; 0 other;

Query Match       100.0%;   Score 6854;   DB 19;   Length 6854;
Best Local Similarity  100.0%;   Pred. No. 0;
Matches 6854;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy       1 CTGCAGGGTTCCCTGGTGCACGACCCGCCCCTGGTCGACGACCAGGGCGCTGTCGCAGAT    60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGCAGGGTTCCCTGGTGCACGACCCGCCCCTGGTCGACGACCAGGGCGCTGTCGCAGAT    60

Qy      61 CGCGGGCGATGTCGGGCGATGTCGGGCGACCACCACGGTGGTGAGCACCACCACGGTGGTG   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGCGGGCGATGTCGGGCGATGTCGGGCGACCACCACGGTGGTGAGCACCACCACGGTGGTG   120

Qy     121 GGCGCGGTTGACCAGCCGGCCACCGCCACCGCGTCCTTCAGCACCAGGCCGATCGTGGG   180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGCGCGGTTGACCAGCCGGCCACCGCCACCGCGTCCTTCAGCACCAGGCCGATCGTGGG   180

Qy     181 CTCGTCCCAGAACACGACACGGCCGGCCGGGTCGTCGTGCAGCAGCCTGCAGCAGCCGGCAT   240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTCGTCCCAGAACACGACACGGCCGGCCGGGTCGTCGTGCAGCAGCCTGCAGCAGCCGGCAT   240

Qy     241 GGCGTCGTCCGAGGCGAGGCTGAGCTGCCCACGGGGGTGACCCCAGCGCGTGGACCCGTCGGG   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCGCTGTCCGAGGCGAGGCTGAGCTGCCCACGGGGGTGACCCCAGCGCGTGGACGTCGGG   300

Qy     301 GTCCCGGAACAGGGCGAGGTTGCGCCCGGTTGCGCCCGGTCCGGGGATGTCGTAGATGCGGGCG   360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTCCCGGAACAGGGCGAGGTTGCGCCCGGTCCGGGGATGTCGTAGATGCGGGCG   360

Qy     361 CAGGATGCGGAAGGAGTCGGGTACCGACAGGTCCCACCAGAGCTGCCTGGCCGAA   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAGGATGCGGAAGGAGTCGGGTACCGACAGGTCCCACCAGAGCTGCCTGGCCGAA   420

Qy     421 GACGACGCCGATCGTGCGGCGGCGGTTGCGCTGCGGTGCCGGTGAGGCTTCCAGCCCGGCGAC   480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GACGACGCCGATCGTGCGGCGGCGGTTGCGCTGCGGTGCCGGTGAGGCTTCCAGCCCGGCGAC   480

Qy     481 CGTGCAGGGCGCCGGAGTGGGGGTCATGATGCCGGTCAGCATCTTGATCGTGGTCGACTT   540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CGTGCAGGGCGCCGGAGTGGGGGTCATGATGCCGGTCAGCATCTTGATCGTGGTCGACTT   540

Qy     541 GCCGGCTCCGTTGGCGCCGATGTAGGCGGTCTTCGTGCCGGCGTATCTCGAAGGAGAC   600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GCCGGCTCCGTTGGCGCCGATGTAGGCGGTCTTCGTGCCGGCGTATCTCGAAGGAGAC   600

Qy     601 GTCGTCGACGGCGCCACGACGCGGTACCGGGGTCAGGAGGGTGGAGAGGCTGCCGAG   660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTCGTCGACGGCGCCACGACGCGGTACCGGGGTCAGGAGGGTGGAGAGGCTGCCGAG   660

Qy     661 CAGGCCGGCTCGCGTTCGGCCACGGCCGTTCCTTGACGAGTGTTCGGCCACGATCAC   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CAGGCCGGCTCGCGTTCGGCCACGGCCGTTCCTTGACGAGTGTTCGGCCACGATCAC   720

Qy     721 GCGATCACCCGCTCGACGGCCGTCTCCAGCAGGCCGAGGCCCTCGTCGAGCAGGCCCTCG   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCGATCACCCGCTCGACGGCCGTCTCCAGCAGGCCGAGGCCCTCGTCGAGCAGGCCCTCG   780

Qy     781 TCGAGGGTGAACGGCGGTGCCAGCCGCAGGATGTGGCCCGCCAGGGAGTGCCAGCCCC   840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TCGAGGGTGAACGGCGGTGCCAGCCGCAGGATGTGGCCCGCCAGGGAGTGCCAGCCCC   840

Qy     841 AGGTCGAGGGCGGCGGTGGTGTAGACGGCCCGGGCGGTGGCGGTGCCGGTGCTCCGGCCGACG   900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AGGTCGAGGGCGGCGGTGGTGTAGACGGCCCGGGCGGTGGCGGTGCCGGTGCTCCGGCCGACG   900

Qy     901 GCGTCGGTGACGAACTCCAGGGCCCCACAGTCCGAGCCCGGTGGCCGGGTACCTGGCCGAGCTGG   960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GCGTCGGTGACGAACTCCAGGGCCCCACAGTCCGAGCCCGGTGGCCGGGTACCTGGCCGAGCTGG   960

Qy     961 GGGAAGCGGGACTCCAGGGCCGCAGCCGCGTCTCTGATGAGCTCCGAGGACGGCACG   1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGGAAGCGGGACTCCAGGGCCGCAGCCGCGTCTCTGATGAGCTCCGAGGACGGCACG   1020

Qy    1021 CGGTCGATCAGCCGGTCGGCGCTCGACGACCTCCAGCGTGGCCGGCGGCGGCGGGATCCCC   1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGGTCGATCAGCCGGTCGGCGCTCGACGACCTCCAGCGTGGCCGGCGGCGGCGGCGGGATCCCC   1080

Qy    1081 AGTGGGTTGCTGCGTACGTGGAGGCGTGACGGGTGCCGCCTCCGGCCTCCGGCCTCGGCA   1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AGTGGGTTGCTGCGTACGTGGAGGCGTGACGGGTGCCGCCTCCGGCCTCCGGCCTCGGCA   1140

Qy    1141 GCTTCCGCGCGTCCGGCCAGCACGGGGGAATCCGCTCGGTGGTGCCCTTGGACAGC   1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GCTTCCGCGCGTCCGGCCAGCACGGGGGAATCCGCTCGGTGGTGCCCTTGGACAGC   1200

Qy    1201 ATCCCAGTGCGCTCGATGCGGATGCCGGATCGTGGG   1260
           ||||||||||||||||||||||||||||||||||||
Db    1201 ATCCCAGTGCGCTCGATGCGGATGCCGGATCGTGGG   1260

Qy    1261 CCGCCGGTGAGGACCTCGTCGGCGACGCAGCAGCAGCGCCGTCCCGCGCCAGGCGCCGGCG   1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CCGCCGGTGAGGACCTCGTCGGCGACGCAGCAGCAGCGCCGTCCCGCGCCAGGCGCCGGCG   1320

Qy    1321 ATCCGCTCCCAGTAGCCGGGGTCGCAGCAGGGGTGACCCCAGCGCGTCGATGTCGAGGAG   1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 ATCCGCTCCCAGTAGCCGGGGTCGCAGCAGGGGTGACCCCAGCGCGTCGATGTCGAGGAG   1380

Qy    1381 AAGACCAGGGCGGAGACGTTGGGCTTCTCCGGATGTCGTAGATGCGGCG   1440
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AAGACCAGGGCGGAGACGTTGGGCTTCTCCGGATGTCGTAGATGCGGCG   1440

Qy    1441 CGCAGTCGCACGAGGGGTACTCCAGGCCCAGGACCCGGTAGCCGTAGGGGCTGTA   1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CGCAGTCGCACGAGGGGTACTCCAGGCCCAGGACCCGGTAGCCGTAGGGGCTGTA   1500

Qy    1501 GCCAGCACGCGTGTTGCGCTGAAGGCCTGGTGGCCGATGTCCAGTGACCAGCCATCCGG   1560

Left column:

```
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 GCCAGCACGTCGTGCCGCTGCCGGCCTGAAGGCCGGATCGCCCAGTGACCGGATCCCCAGT 1560

Qy 1561 GCGCCCATGGTCGTTGCCGTGGCGCAGGGCGCCGAGATCCGGTTGCGGCCCGGC 1620
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 GCGCCCATGGTCGTTGCCGTGGCGCAGGGCGCCGAGATCCGGTTGCGGCCCGGC 1620

Qy 1621 GCGGCGGTCGCGCCTGGACGACCCCGCAGGAGGCGCCTCGACCACCTCCGGCGCCGGTGAGAAG 1680
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 GCGGCGGTCGCGCCTGGACGACCCCGCAGGAGGCGCCTCGACCACCTCCGGCGCCGGTGAGAAG 1680

Qy 1681 AAGGCGTAGGTGTCGAGCTGTTCGGCAGCCTGGGCGCAGCAGTTCCAGCAGGCCGGCG 1740
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 AAGGCGTAGGTGTCGAGCTGTTCGGCAGCCTGGGCGCAGCAGTTCCAGCAGGCCGGCG 1740

Qy 1741 CGGTCCGGCGGCGTGTCGTCAGGGCGTGCCCGGTGACGTCGTGAGGCGTCCCGGAAGTCG 1800
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 CGGTCCGGCGGCGTGTCGTCAGGGCGTGCCCGGTGACGTCGTGAGGCGTCCCGGAAGTCG 1800

Qy 1801 TCGACGAACTCCGGGCTGCCCGGTGACGGGTGAGGGTCCGGCGCCGCGAAGTCG 1860
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1801 TCGACGAACTCCGGGCTGCCCGGTGACGGGTGAGGGTCCGGCGCCGCGAAGTCG 1860

Qy 1861 AGTGACTGGTTGCCGTCCAGGTCGGTGGCGACGGGACCCGTCCCTCGGCGCGAAGACCCGG 1920
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1861 AGTGACTGGTTGCCGTCCAGGTCGGTGGCGACGGGACCCGTCCCTCGGCGCGAAGACCCGG 1920

Qy 1921 CGTCCGTGACGGGCTTCCTCGGAGGCGCCCGGCGGCCTCCCGTGCCAGG 1980
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1921 CGTCCGTGACGGGCTTCCTCGGAGGCGCCCGGCGGCCTCCCGTGCCAGG 1980

Qy 1981 TGCGTGTCTCGCCGTAAGCCTGTCATCGCGTCGCTGCCTCGTCGACCGGCTGACGCGAT 2040
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1981 TGCGTGTCTCGCCGTAAGCCTGTCATCGCGTCGCTGCCTCGTCGACCGGCTGACGCGAT 2040

Qy 2041 CGCCGGCGAACTGCCGTTGTGGCGCCACCACGGTTGGGGCGGCTCGAGTCAAACAC 2100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2041 CGCCGGCGAACTGCCGTTGTGGCGCCACCACGGTTGGGGCGGCTCGAGTCAAACAC 2100

Qy 2101 TTGAACACACACCGCTCGGTCAAGAGTTGCGGGTTGTTTCAGAAAGTTGTTGCGAGCCGGAT 2160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2101 TTGAACACACACCGCTCGGTCAAGAGTTGCGGGTTGTTTCAGAAAGTTGTTGCGAGCCGGAT 2160

Qy 2161 CGGCACTCTGGTTGAGTGCACGTGCTTACGGCGCCACCACGGTTCGAGGAGGGGA 2220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2161 CGGCACTCTGGTTGAGTGCACGTGCTTACGGCGCCACCACGGTTCGAGGAGGGGA 2220

Qy 2221 CCTGTGAGAACAAGCCCGCAGACCCGGTACGGCGCTCCCGGAGGCCCTGG 2280
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2221 CCTGTGAGAACAAGCCCGCAGACCCGGTACGGCGCTCCCGGAGGCCCTGG 2280

Qy 2281 TCCTGGCAGGTGAACCGGCAGCAGACTGAGGCCGTTCACCCACCACCGCCAAGCAGC 2340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 TCCTGGCAGGTGAACCGGCAGCAGACTGAGGCCGTTCACCCACCACCGCCAAGCAGC 2340

Qy 2341 TGCTCCCATCGCCAACAGCCCGGCGTCTTCTACGCGTCTGGAGTCCCGCCGCGCGG 2400
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 TGCTCCCATCGCCAACAGCCCGGCGTCTTCTACGCGTCTGGAGTCCCGCCGCGCGG 2400

Qy 2401 GTGTCCGGGAGGCCGGCCGTCGTCTGGGCGTACGGCGGCCGAGATCCGGCGAACTCACCG 2460
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 GTGTCCGGGAGGCCGGCCGTCGTCTGGGCGTACGGCGGCCGAGATCCGGCGAACTCACCG 2460

Qy 2461 GCGACGGCACCGCGGTTCGGGTTACGCATCACCTACCTCCGGGGCCGTTCCTGGGTCGTC 2520
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2461 GCGACGGCACCGCGGTTCGGGTTACGCATCACCTACCTCCGGGGCCGTTCCTGGGTCGGTC 2520

Qy 2521 TCGGGCCACCCGGTTCGGCGTGCGCATCGCGCCGGCGACGACTTCCTGCTGTACC 2580
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 TCGGGCCACCCGGTTCGGCGTGCGCATCGCGCCGGCGACGACTTCCTGCTGTAACC 2580

Qy 2581 TGGGGCACAACTACCTGCCCCAGGGGGTCACCCGCCGCCAATCGCGCCCGATC 2640
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Right column:

```
Db 2581 TGGGGCACAACTACCTGCCCCAGGGGGTCACCCGACTTCGCCCGCGTCACCGCCGCGCCGATC 2640

Qy 2641 CCGCGGCGGCCCGGCGTCGCTGCTGCCGGTTGCGGCCCGTCGCGTCGCGG 2700
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 CCGCGGCGGCCCGGCGTCGCTGCTGCCGGTCGCTGCCGTCGCGTCGCGG 2700

Qy 2701 AGGTCGACGCGGACGGGAACGTGCTGCGCTTGGAGGAGAAACCCGACGTCCCGCGCAGCT 2760
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2701 AGGTCGACGCGGACGGGAACGTGCTGCGCTTGCGCTTGGAGGAGAAACCCGACGTCCCGCAGCT 2760

Qy 2761 CGCTCCGCTCATCGGCGTGTACGCGTTCAGCCGCCGTCAGCAGGCGGTACGGGCCA 2820
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2761 CGCTCCGCTCATCGGCGTGTACGCGTTCAGCCGCCGTCAGCAGGCGGTACGGGCCA 2820

Qy 2821 TCACCCCCTCCGCGCCCGGCGGGCCGTACGGGCGTGGAGATCACCCACCGCCGTCAGTGGATGATCGACC 2880
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 TCACCCCCTCCGCGCCCGGCGGGCCGTACGGGCGTGGAGATCACCCACCGCCGTCAGTGGATGATCGACC 2880

Qy 2881 GGGGCCTGCGCGGTACGGGCCCGGAGACCACCACCCGGCCCTGCGCGCCGACCACCGCAGCCGCGG 2940
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 GGGGCCTGCGCGGTACGGGCCCGGAGACCACCACCCGGCCCTGCGCGCCGACCACCGCAGCCGCGG 2940

Qy 2941 AGGACATGCTGGAGGTCAACCGTCACGTCCTGACGACTGCGAGGGGA 3000
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 AGGACATGCTGGAGGTCAACCGTCACGTCCTGACGACTGAGGGGCCGATCGAGGGGA 3000

Qy 3001 AGGTCGACGCGCACAGCAGCGTCGGTCCGGTGCCGCGGGTCCGGAAGGCGCGATCGTCTCA 3060
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3001 AGGTCGACGCGCACAGCAGCGTCGGTCCGGTGCCGCGGGTCCGGAAGGCGCGATCGTGC 3060

Qy 3061 GGGGGTCACACGTGGTTGGCGCCGGGTGTGATCGGCGGTGCCGTCGTCGTCAGCAACTCCA 3120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3061 GGGGGTCACACGTGGTTGGCGCCGGGTGTGATCGGCGGTGCCGTCGTCGTCAGCAACTCCA 3120

Qy 3121 GTGTCGGCCCGTACACCTCCATCGGGAGGACGTCAGTCAAACAC 3180
        ||||||||||||||||||||||||||||||||||||||||||||
Db 3121 GTGTCGGCCCGTACACCTCACCTCCATCGGGAGGACGTCGAGGAGT 3180

Qy 3181 ACTCCGTCCTGTCTGCGCGGCCCGCCCAGTCGAGGGGCATCGAGGGCTCCCTCA 3240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3181 ACTCCGTCCTGTCTGCGCGGCCCGCCCAGTCGAGGGGCATCGAGGGCTCCCTCA 3240

Qy 3241 TCGGCCGCGGCGCCGTCGTGCGTGTGGCCCGCAGGTCACCGACTGGTGA 3300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3241 TCGGCCGCGGCGCCGTCGTGCGTGTGGCCCGCAGGTCACCGACTGGTGA 3300

Qy 3301 TCGGCGACCAGCAGGTGTATCTCCGCCTCCCAGTGTCCCATGACCAACGACCATCCTCGTCACCGGCGG 3360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3301 TCGGCGACCAGCAGGTGTATCTCCGCCTCCCAGTGTCCCATGACCAACGACCATCCTCGTCACCGGCGG 3360

Qy 3361 AGCGGGCTTCATTCGCTCCGCCTACGTCGTGTCGTCTCGTGCCCGGGGCGGG 3420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3361 AGCGGGCTTCATTCGCTCCGCCTACGTCGTGTCGTCTCGTGCCCGGGGCGGG 3420

Qy 3421 CGTCGGCGGTGACCGTCCTCGACAAACTCACCTACGCGCAGCCTCGCCCGCTGCACGC 3480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3421 CGTCGGCGGTGACCGTCCTCGACAAACTCACCTACGCGCAGCCTCGCCCGCTGCACGC 3480

Qy 3481 GGTGCCTGACCATCCGCCTCACCTTCGTCAAGGTGTGTGCCGAGTGTGGCCACCGGCGTCGT 3540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3481 GGTGCCTGACCATCCGCCTCACCTTCGTCAAGGTGTGTGCCGAGTGTGGCCACCGGCGTCGT 3540

Qy 3541 CGACACGCTGGCGCGCGCAGCGGTGCCTTCGGCGCGACATCGTGCACTTCGGCGCGACGTCGTC 3600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3541 CGACACGCTGGCGCGCGCAGCGGTGCCTTCGGCGCGACATCGTGCACTTCGGCGCGACGTCGA 3600

Qy 3601 CCGCTCCATCACCGACACGCGGCCCTCCGGGGACGACGACTTCCTGCTGTACC 3660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3601 CCGCTCCATCACCGACACGCGGCCCTCCGGGGACGACGACTTCCTGCTGTACC 3660

Qy 3661 GCTCGACGCGCCCGGCTCCGCCACCTTCGTGCGCACCGTGTCCACCGACGAGGT 3720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3661 GCTCGACGCGCCCGGCTCCGCCACCTTCGTGCGCACCGTGTCCACCGACGAGGT 3720
```

Left column:

```
Qy  3721 GTACGGCTCCCTCCCGCACGGGGCGCGGGAGAGCGACCCCTGCTTCGGACCTCGCC 3780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721 GTACGGCGGCGTCCGGGTGACCCGCGTCGGGGAGAGCGACCCCTGCTTCGGACCTCGCC 3780

Qy  3781 GTACGGCGGCGTCGAAGGCGGCCTCGGACCTCATGGCGCTCGGGGAGAGCGACCCCTGGAA 3840
         ||||||||||||||||||||||||||||||||||||
Db  3781 GTACGGCGGCGTCGAAGGCGGCCTCGGACCTCATGGCGCTCGGGGAGAGCGACCCCTGGAA 3840

Qy  3841 CCTGGACGTCCGGGTGACCCGCGTGTTCGAACAACTTCGGCCCCCACCAGCATCCGAGAA 3900
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  3841 CCTGGACGTCCGGGTGACCCGCGTGTTCGAACAACTTCGGCCCCCACCAGCATCCGAGAA 3900

Qy  3901 GCTCATACCGGCGGCTTCCTGACCAGCCTCGTGTCCGGCGGCACCGTTCCCCTCTACGGCGA 3960
         ||||||||||||||||||||||||||||||||||||||
Db  3901 GCTCATACCGGCGGCTTCCTGACCAGCCTCGTGTCCGGCGGCACCGTTCCCCTCTACGGCGA 3960

Qy  3961 CGGGCGGCACGTGCGCGACGTCGCTGCTCGACGACCACGTCAGGGCGCCGTCGAACTCGT 4020
         ||||||||||||||||||||||||||||||||||||||
Db  3961 CGGGCGGCACGTGCGCGACGTCGCTGCTCGACGACCACGTCAGGGCGCCGTCGAACTCGT 4020

Qy  4021 CCGCGTGTCGGGGAGAGATCTACAACATCGGGGCGGCGGCCCCGCACCTTCGCTCGCTCCCAA 4080
         ||||||||||||||||||||||||||||||||||||
Db  4021 CCGCGTGTCGGGGAGAGATCTACAACATCGGGGCGGCGGCCCCGCACCTTCGCTCGCTCCCAA 4080

Qy  4081 CCTGGAGCGTCACGCACCGTTGCTCGCCACTGTGCGGCGGCGCCCGAGCGCATCGTCCA 4140
         |||||||||||||||||||||||||||||||||||||||
Db  4081 CCTGGAGCGTCACGCACCGTTGCTCGCCACTGTGCGGCGGCGCCCGAGCGCATCGTCCA 4140

Qy  4141 CGTCGAGAACCGCAAGGGCACGACCGGCGCTACGCGGTCGGTCACGGGCGTCGACCAAGAATCACCGGC 4200
         |||||||||||||||||||||||||||||||||
Db  4141 CGTCGAGAACCGCAAGGGCACGACCGGCGCTACGCGGTCGGTCACGGGCGTCGACCAAGAATCACCGGC 4200

Qy  4201 GGAACTCGGTTACCGGCGCGCCGCGCCGACTTCGGCGGCCAGTTCGGCGGCCTGGCCCCGACCACC 4260
         ||||||||||||||||||||||||
Db  4201 GGAACTCGGTTACCGGCGCGCCGCGCCGACTTCGGCGGCCAGTTCGGCGGCCTGGCCCCGACCACC 4260

Qy  4261 GTACGAGCGGCGCAGAGGAGACTGGTCGGTCCCCTGCTCGCCGCCGCGCCGCCGCCGCGG 4320
         |||||||||||||||||||||||||||||
Db  4261 GTACGAGCGGCGCAGGAGACTGGTCGGTCCCCTGCTCGCCGCCGCGCCGCCGCCGCGG 4320

Qy  4321 ACCGCAACCACCGGCCCCGGCCCCGGCGCACACCCCGGCGGCCGCCGCGGCGGCGCGG 4380
         |||||||||||||||||||||||||||||||||
Db  4321 ACCGCAACCACCGGCCCCGGCCCCGGCGCACACCCCGGCGGCCGCCGCGGCGGCGGCGG 4380

Qy  4381 CGTCCGTGAGCGGAGCGGCCATGAGCGGCGAGTCGTCGAAGCGGCCACCCCGGACCACC 4440
         |||||||||||||||||||||||||||||||||
Db  4381 CGTCCGTGAGCGGAGCGGCCATGAGCGGCGAGTCGTCGAAGCGGCCACCCCGGACCACC 4440

Qy  4441 AGTTCCGGCATGAAGAGACGAATTCGGTGCGCGGCGGCGGCCGGCGTTCGGTGCGCGGCGGCGGCGCGG 4500
         |||||||||||||||||||||||
Db  4441 AGTTCCGGCATGAAGAGACGAATTCGGTGCGCGGCGGCGGCCGGCGTTCGGTGCGCGGCGGCGGCGCGG 4500

Qy  4501 AGTGCGTCCACGGCGACCTGCCTTGACGGGCTGTCTCGATGGTGGTCTCGATGGTGGTCAGGGGA 4560
         ||||||||||||||||||||||||||
Db  4501 AGTGCGTCCACGGCGACCTGCCTTGACGGGCTGTCTCGATGGTGGTCTCGATGGTGGTCAGGGGA 4560

Qy  4561 GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCGCCACACCGGCCATCATGTCGCCTGGGCG 4620
         ||||||||||||||||||||||||||||||||
Db  4561 GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCGCCACACCGGCCATCATGTCGCCTGGGCG 4620

Qy  4621 ACCGTGAGACGCCCGCCGGCGGCGCCGCGGACGCGGCGGCGCGGACGCGGCCGCGGCGGCG 4680
         ||||||||||||||||||||||||||||||||||
Db  4621 ACCGTGAGACGCCCGCCGGCGGCGCCGCGGACGCGGCGGCGCGGACGCGGCCGCGGCGGCG 4680

Qy  4681 CACATGACGGCGGTCAGCGATCAGCGCTGCGATCAGCGGCGGCGCGACTCCGCGACACTCCCAGGTGC 4740
         |||||||||||||||||||||||||||||
Db  4681 CACATGACGGCGGTCAGCGATCAGCGCTGCGATCAGCGGCGGCGCGACTCCGCGACACTCCCAGGTGC 4740

Qy  4741 AGGGAGAAACAGCGAGTGCTGCACGAGTCGCTGCCGACACTCCGCGACACTCCCAGGTGC 4800
         ||||||||||||||||||||||||||||||||
Db  4741 AGGGAGAAACAGCGAGTGCTGCACGAGTCGCTGCCGACACTCCGCGACACTCCCAGGTGC 4800
```

Right column:

```
Qy  4801 TCCCGCACGCGGCCGGCGGCCCGGAACCCCTGCTGCACGCGGGCACGAGCGGCGGGGC 4860
         |||||||||||||||||||||||||||||||||||||
Db  4801 TCCCGCACGCGGCCGGCGGCCCGGAACCCCTGCTGCACGCGGGCACGAGCGGCGGGGC 4860

Qy  4861 CCGACGGCGAGGCCGACGCGCTCGTCGTCGCTCCGCCAGTGCGGCACGGCGCCCAGGCGC 4920
         ||||||||||||||||||||||||||||||||||
Db  4861 CCGACGGCGAGGCCGACGCGCTCGTCGTCGCTCCGCCAGTGCGGCACGGCGCCCAGGCGC 4920

Qy  4921 ATCGCGGCCGCCGGCCGTCGTCCGGGGAGACGAAGGTGCCTCGATCCGGGGCGAGAACCGTTC 4980
         |||||||||||||||||||||||||||||
Db  4921 ATCGCGGCCGCCGGCCGTCGTCCGGGGAGACGAAGGTGCCTCGATCCGGGGCGAGAACCGTTC 4980

Qy  4981 ACGAGGACGAAGGGCACCTGCCGCTCGTGAGCGGCGCCGTACCGTCGGGTCCGGCGGTG 5040
         ||||||||||||||||||||||||||||||||
Db  4981 ACGAGGACGAAGGGCACCTGCCGCTCGTGAGCGGCGCCGTACCGTCGGGTCGCGCGGTG 5040

Qy  5041 GTGTCCGCGTGCAGTTCCGAGACGAAGATGATGCCGGACACACCCGGTCCACGAGCATC 5100
         |||||||||||||||||||||||||||||||||
Db  5041 GTGTCCGCGTGCAGTTCCGAGACGAAGATGATGCCGGACACACCCGGTCCACGAGCATC 5100

Qy  5101 TCCGTGAGTTCGTTCCTCCGTCGTGGTCGCAGCCGTCGCGTGGCGAGCACGGGCGTGTAG 5160
         ||||||||||||||||||||||||||||||
Db  5101 TCCGTGAGTTCGTTCCTCCGTCGTGGTCGCAGCCGTCGCGTGGCGAGCACGGGCGTGTAG 5160

Qy  5161 CCCTGACGCGTGAGCGCCTGCGCCCATCACCTGGCCCAGTGCGGGGAAGAAGGGGTTGTCC 5220
         ||||||||||||||||||||||||||||||||
Db  5161 CCCTGACGCGTGAGCGCCTGCGCCCATCACCTGGCCCAGTGCGGGGAAGAAGGGGTTGTCC 5220

Qy  5221 AGTTCGGGGGTGACCAGTCCGACCAGTCGCGCGTGCGCCGTCTGCGCGCGCCGCA 5280
         |||||||||||||||||||||||||||||
Db  5221 AGTTCGGGGGTGACCAGTCCGACCAGTCGCGCGTGCGCCGTCTGCGCGCGCCGCA 5280

Qy  5281 CCCAGCGCGTCCAGTGCGTTCAGCGCGACGTGCGGGTGCCCGGTGCCGGTGCCCACCGGCGCCGCCGCA 5340
         ||||||||||||||||||||||||||||||||
Db  5281 CCCAGCGCGTCCAGTGCGTTCAGCGCGACGTGCGGGTGCCCGGTGCCGGTGCCCACCGGCGCCGCCGCA 5340

Qy  5341 CCGTTCAGCACCACCCGGCTGACCGTGCCCTTGCTGCGACGCCACTCTACCTGTCGGGGCGTGCCCACCCGCGC 5400
         ||||||||||||||||||||||||||||
Db  5341 CCGTTCAGCACCACCCGGCTGACCGTGCCCTTGCTGCGACGCCACTCTACCTGTCGGGGCGTGCCCACCCGCGC 5400

Qy  5401 AGCCGCATGGTCATGGCAACGCACTCTACCTGTCGGGGCGTGCCCACCGGCCGCCCCGAAA 5460
         ||||||||||||||||||||||||||||||||
Db  5401 AGCCGCATGGTCATGGCAACGCACTCTACCTGTCGGGGCGTGCCCACCGGCCGCCCCGAAA 5460

Qy  5461 GCGGAACCGGCGGACTGCGGGACTGCGGGGCACGGCCACGCGGACCACGCGCCCGAAA 5520
         |||||||||||||||||||||||||||||||||
Db  5461 GCGGAACCGGCGGACTGCGGGACTGCGGGGCACGGCCACGCGGACCACGCGCCCGAAA 5520

Qy  5521 CGATGGCTGAAAATGCTTGCAGCAAATGCCGCGCGCGCAGAAGCGGTTGCGGTGAGGCGTTTCGGATCCT 5580
         |||||||||||||||||||||||||||||
Db  5521 CGATGGCTGAAAATGCTTGCAGCAAATGCCGCGCGCGCAGAAGCGGTTGCGGTGAGGCGTTTCGGATCCT 5580

Qy  5581 GTTACGTTCCTGCAACCCGCGCCGCGCCAGAGGAGTTCACAATGCGGCGTGCATTGCGGCGTGCCACCG 5640
         |||||||||||||||||||||||||||
Db  5581 GTTACGTTCCTGCAACCCGCGCCGCGCCAGAGGAGTTCACAATGCGGCGTGCATTGCGGCGTGCCACCG 5640

Qy  5641 CCGCCCGATTCCGGGGATCACTCAGGGAGTTCACAATGCGGCGTGCATTGCGGCGTGCCACCG 5700
         |||||||||||||||||||||||||||||||
Db  5641 CCGCCCGATTCCGGGGATCACTCAGGGAGTTCACAATGCGGCGTGCATTGCGGCGTGCCACCG 5700

Qy  5701 CGCTGTTCGCGGCGTGTGGCCATGACGGCCATGACGGCGCGCACGCGTCGGGGCACCTTCGGGACACGTCCA 5760
         ||||||||||||||||||||||||||||||
Db  5701 CGCTGTTCGCGGCGTGTGGCCATGACGGCCATGACGGCGCGCACGCGTCGGGGCACCTTCGGGACACGTCCA 5760

Qy  5761 GCGGTACGACGCGGCGGCCACGTACCAGGGCCTCGCGGAGGAGCGCTTCGAGAAGGGAGCCACCCCGA 5820
         ||||||||||||||||||||||||||||||||
Db  5761 GCGGTACGACGCGGCGGCCACGTACCAGGGCCTCGCGGAGGAGCGCTTCGAGAAGGGAGCCACCCCGA 5820

Qy  5821 ACGAAGCCGAGAAGGCGACGTACCAGGGCCTCGCGGAGGAGCGCTTCGAGAAGGGAGCCACCCCGA 5880
         ||||||||||||||||||||||||||||
Db  5821 ACGAAGCCGAGAAGGCGACGTACCAGGGCCTCGCGGAGGAGCGCTTCGAGAAGGGAGCCACCCCGA 5880

Qy  5881 AGGTCGACGTCAAGTACGTCAACGTCCCGTTCGGCGAGGCGAGCGAACGCCAAGTTCAAGAACG 5940
```

```
XX
DT   26-SEP-2001 (first entry)
XX
DE   Micromonospora DNA encoding biosynthetic enzymes for Everinomycin.
XX
KW   Everinomicin; antibiotic; bottle-neck gene; orthomicin;
KW   fermentation; ds.
XX
OS   Micromonospora carbonacea var. africana.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (132..1382)
FT                   /*tag= a
FT                   /product= "EvdA"
FT   RBS             complement (1389..1394)
FT                   /*tag= b
FT   CDS             complement (1490..2611)
FT                   /*tag= c
FT                   /product= "EvdB"
FT   RBS             complement (2618..2622)
FT                   /*tag= d
FT   CDS             complement (2622..3860)
FT                   /*tag= e
FT                   /product= "EvdC"
FT   RBS             complement (3867..3870)
FT                   /*tag= f
FT   CDS             4143..5312
FT                   /*tag= g
FT                   /product= "EvdD"
FT   RBS             4134..4138
FT                   /*tag= h
FT   CDS             5309..6235
FT                   /*tag= i
FT                   /product= "EvdE"
FT                   6232..7275
FT                   /*tag= j
FT                   /product= "EvdF"
FT   RBS             6226..6229
FT                   /*tag= k
FT   CDS             7272..8327
FT                   /*tag= l
FT                   /product= "EvdG"
FT                   8342..9364
FT                   /*tag= m
FT   RBS             /product= "EvdH"
FT                   8333..8336
FT                   /*tag= n
FT   CDS             complement (9463..10224)
FT                   /*tag= o
FT                   /product= "EvdI"
FT   RBS             complement (10232..10235)
FT                   /*tag= p
FT   CDS             10424..11176
FT                   /*tag= q
FT                   /product= "EvdJ"
FT   CDS             12027..12455
FT                   /*tag= r
FT                   /product= "EvdK"
FT                   /partial
FT                   /note= "No start codon"
FT   CDS             complement (12108..13022)
FT                   /*tag= s
FT                   /product= "EvdL"
FT   RBS             complement (13027..13030)
FT                   /*tag= t
FT   CDS             complement (14410..15363)
FT                   /*tag= u
FT                   /product= "EvrA"
FT   RBS             complement (15369..15373)
FT                   /*tag= v
FT   CDS             complement (15380..16414)
FT                   /*tag= w
FT                   /product= "EvrB"
```

```
Db    5881  |||||||||||||||||||||||||||||||||||||||||||
            AGGTCGACGTCAAGTACGTCCCGTTCGGCGGAGGCGAACGTTCAAGAACG  5940
Qy    5941  CCGGGGGCGGCAACTCCGGTGCCCCGGACGTGATGCGTACGGAGCTCGCCGG  6000
Db    5941  CCGGGGGCGGCAACTCCGGTGCCCCGGACGTGATCGCTACGGAGGTCGCCGG  6000
Qy    6001  ACTTCGCCAGCATCGGCTACCTCGCCCGCTCGACGGCACGCCCGCCCTCGACGACGACGGGT  6060
Db    6001  ACTTCGCCAGCATCGGCTACCTCGCCCGCTCGACGGCACGCCCGCCCTCGACGACGCGGGT  6060
Qy    6061  CGGACCACCTTCCCCAGGGCGGCAGCACCAGGTACGAGGGGAAGACCTACGCGGTCCCGC  6120
Db    6061  CGGACCACCTTCCCCAGGGCGGCAGCACCAGGTACGAGGGGAAGACCTACGCCGG'TCCCGC  6120
Qy    6121  AGGTGATCGACACCCTGGCGCTCTTCTACAACAAGGAACTGCTGACGAAGGCCGTGTCG  6180
Db    6121  AGGTGATCGACACCCTGGCGCTCTTCTACACACAAGGAACTGCTGACGAAGGCCGTGTCG  6180
Qy    6181  AGGTGCCGGGCTCCCCGCGAGTGAAGACGTGAAGACGGCGCCGCCGAGATCACCGAGAAGACCG  6240
Db    6181  AGGTGCCGGGCTCCCCGCGAGTGAAGACGTGAAGACGGCGCCGCCGAGATCACCGAGAAGACCG  6240
Qy    6241  GCGCCAGCGGCCGCCTTCACTGCGGGGGCGACGACCCCGTACTTGGTTCCTGCCCTACCTCTACG  6300
Db    6241  GCGCCAGCGGCCGCCCTTCACTGCGGGGGCGACGACCCCGTACTTGGTTCCTGCCCTACCTCTACG  6300
Qy    6301  GGGAGGGGCGGCGGCGACCTGGTCGACAGAGAACAAGACCGTCACGCGTCGACGACGAAGCCG  6360
Db    6301  GGGAGGGGCGGCGGCGACCTGGTCGACAGAGAACAAGACCGTCACGCGTCGACGACGAAGCCG  6360
Qy    6361  GTGTGCGCGGCCTACCGCGTCATCAAGGACCTCGTGACGACCAAGGCGGCCATCACCGACG  6420
Db    6361  GTGTGCGCGGCCTACCGCGTCATCAAGGACCTCGTGACGACCAAGGCGGCCATCACCGACG  6420
Qy    6421  CGTCCGACGGCTGGAACAACATGCAGAACGCCTTCAAGTCGGGCCAAGGTCGCCATGATGG  6480
Db    6421  CGTCCGACGGCTGGAACAACATGCAGAACGCCTTCAAGTCGGGCCAAGGTCGCCATGATGG  6480
Qy    6481  TCAACGGCCCCTGGGCCATCGAGGACGTCAAGGCGGGGAGCCCGCTTCAAGGACGCCGGCA  6540
Db    6481  TCAACGGCCCCTGGGCCATCGAGGACGTCAAGGCGGGGAGCCCGCTTCAAGGACGCCGGCA  6540
Qy    6541  ACCTGGGGGTCGCCCCCGCCCGGTCGCCGGCAGTCCGGCGCTCGACGCGGCGGGT  6600
Db    6541  ACCTGGGGGTCGCCCCCGCCCGGTCGCCGGCAGTGCCGGCGCTCTCCCCAGGGCGGGT  6600
Qy    6601  GGAACCTTCGGTGTACGGCGGGCTTCGAAGAACTCGACGCCTTCGTGAAGT  6660
Db    6601  GGAACCTTCGGTGTACGGCGGGCTTCGAAGAACTCGACGCCTTCGTGAAGT  6660
Qy    6661  ACATGAGCTCCGCCAAGGTGCAGCAGCACCGAGAGCTGAGCCTGCTGCCCACCC  6720
Db    6661  ACATGAGCTCCGCCAAGGTGCAGCAGCACCGAGAGCTGAGCCTGCTGCCCACCC  6720
Qy    6721  GCACGTCCGTCTACGAGGTCCCGTCCGCGGACGATGTGTGAAGTTCTTCAAGC  6780
Db    6721  GCACGTCCGTCTACGAGGTCCCGTCCGCGGACAACGATGTGTGAAGTTCTTCAAGC  6780
Qy    6781  CGGCCGTCGACAAGGCCGTCGAACGGCCGGATCGCCGAGC  6840
Db    6781  CGGCCGTCGACAAGGCCGTCGAACGGCCGGATCGCCGAGC  6840
Qy    6841  CGATCCGGCTGCAG  6854
Db    6841  CGATCCGGCTGCAG  6854
```

```
RESULT 2
AAS08693/c
ID   AAS08693 standard; DNA; 109519 BP.
XX
AC   AAS08693;
```

us-09-922-683-7.rng

```
FT   CDS           complement 16419..17873
FT                 /tag= x
FT                 /product= "EvrC"
FT   CDS           complement (17870..18934)
FT                 /tag= y
FT                 /product= "EvrD"
FT   CDS           19374..20906
FT                 /tag= z
FT                 /product= "EvrE"
FT   CDS           21064..22542
FT                 /tag= aa
FT                 /product= "EvrF"
FT   RBS           21056..22542
FT                 /tag= ab
FT   CDS           22748..24172
FT                 /tag= ac
FT                 /product= "EvrG"
FT   RBS           22736..22740
FT                 /tag= ad
FT   CDS           complement (24177..25223)
FT                 /tag= ae
FT                 /product= "EvrH"
FT   RBS           complement (25230..25233)
FT                 /tag= af
FT   CDS           25550..26626
FT                 /tag= ag
FT                 /product= "EvrI"
FT   RBS           26685..30479
FT                 /tag= ah
FT   CDS           26672..26676
FT                 /tag= ai
FT                 /product= "EvrJ"
FT   RBS           complement (30557..31876)
FT                 /tag= aj
FT   CDS           complement (31885..31888)
FT                 /tag= ak
FT                 /product= "EvrK"
FT   RBS           complement (31941..32882)
FT                 /tag= al
FT   CDS           complement (33167..34405)
FT                 /tag= am
FT                 /product= "EvrL"
FT   CDS           complement (34414..34418)
FT                 /tag= an
FT                 /product= "EvrM"
FT   RBS           complement (34449..35210)
FT                 /tag= ao
FT   CDS           complement (35219..35221)
FT                 /tag= ap
FT                 /product= "EvrN"
FT   RBS           complement (35294..36238)
FT                 /tag= aq
FT   CDS           complement (36235..36963)
FT                 /product= "EvrO"
FT                 /tag= ar
FT   CDS           complement (36998..38026)
FT                 /product= "EvrP"
FT                 /tag= as
FT   CDS           complement (38072..38566)
FT                 /product= "EvrQ"
FT                 /tag= at
FT   CDS           complement (38892..40163)
FT                 /product= "EvrR"
FT                 /tag= au
FT   CDS           complement (40216..40890)
FT                 /product= "EvrS"
FT                 /tag= av
FT   RBS           complement (40899..40902)
FT                 /product= "EvrT"
FT                 /tag= aw
FT   CDS           complement (40887..41576)
FT                 /tag= ax
FT                 /product= "EvrU"

FT   CDS           complement (41679..42707)
FT                 /tag= ay
FT                 /product= "EvrV"
FT   RBS           complement (42714..42717)
FT                 /tag= az
FT   CDS           complement (42810..43799)
FT                 /tag= ba
FT                 /product= "EvrW"
FT   RBS           complement (43807..43811)
FT                 /tag= bb
FT   CDS           complement (43799..44866)
FT                 /tag= bc
FT                 /product= "EvrX"
FT   CDS           complement (45014..45760)
FT                 /tag= bd
FT                 /product= "EvrY"
FT   RBS           complement (45767..45770)
FT                 /tag= be
FT   CDS           complement (45962..46714)
FT                 /tag= bf
FT                 /product= "EvrZ"
FT   RBS           complement (45952..45956)
FT                 /tag= bg
FT   CDS           complement (47156..49234)
FT                 /tag= bh
FT                 /product= "EvsA"
FT   CDS           51627..52715
FT                 /tag= bi
FT                 /product= "EvsB"
FT   RBS           51629..51622
FT                 /tag= bj
FT   CDS           52889..53557
FT                 /tag= bk
FT                 /product= "EvsC"
FT   CDS           53554..54207
FT                 /tag= bl
FT                 /product= "EvbA"
FT   CDS           complement (54362..55117)
FT                 /tag= bm
FT                 /product= "EvbB"
FT   CDS           complement (55125..55128)
FT                 /tag= bn
FT   RBS           complement (55135..56094)
FT                 /tag= bo
FT                 /product= "EvbC"
FT   CDS           complement (56100..56103)
FT                 /tag= bp
FT   RBS           complement (56184..56813)
FT                 /tag= bq
FT   CDS           56961..58709
FT                 /product= "Evbc2"
```

```
Query Match     12.8%;    Score 877.8;    DB 22;    Length 109519;
Best Local Similarity   56.5%;    Pred. No. 2e-100;
Matches 1892;    Conservative   0;    Mismatches 1402;    Indels  53;    Gaps  12:

Qy    961 GGGAAGCGGGACTCCAGGGCGCGCAGCCCGCTCCTGATGAGCTCGCCGAGGACGCGCCACG 1020
          |  |||| ||   |   ||  ||||  ||||||  ||| ||||| || ||   | |||
Db  46149 GCGAACTGGATGCCCGCCGCCTTGCGCGCTCGTCCAGTTGATCGATCCCTGGTAGTGGGAC 46090

Qy   1021 CGGTCGATCAGCCGGTCGCGGGGTCGCGGACCTCCAGCGTGGCCGGCCGCGGCGATCCCC 1080
          ||||||| || ||  |||  ||| ||||||  | ||||||| |||| |||| | |||| |
Db  46089 ACGTCGATGCCGGGGTGGTGCGGCGGCCGCGCCGGGCGACGAGGGTCGCG 46030

Qy   1081 AGTGGGTTGCTCGCGTACGTCGAGGCGTACGCCGGGTGGCCGCCTCCCGCCTCCGCCA 1140
          | | ||| |  ||   |   || | | || | |||  |  |||| |  |||| || |||
Db  46029 GCAGCGGTGGCGAGGCTGGCCAGACCGGCGCAGGGCGGTTCGGGTA 45970

Qy   1141 GCTTCCGCGGCGTCCGGCCCAGCACGGC----GAAGGGGAATCCGCCGCGTGCCCTTGA 1196
          |||| | ||  ||| ||   | || | ||       ||  ||||  | || |||
Db  45969 CGGGTCACGGATGCCTCCAGGATCGCTGAAGAATTGACAGCCATCTTTGAAGAAATCCGC 45910

Qy   1197 CAGCATCGCCAGGTCGCCGGCCTCGATGCAGCAGTTCGCTGGCGAGGAGGCCGGTCGCG 1256
```

```
          |   |  || ||      |   |   |     || ||     | ||     || || |||
Db  45909 CTAGAAGGCAAGAGTGACCGAAAATGCCTGTCGTCCGGACCGGCGGCCATGGCGGGCGCG 45850

Qy   1257 CCCGCCGCCGGTGAGGACCTCGTCGGCGACGAGCAGCACGCCGCCGTCCCGGCAGGCGCC 1316
          |||| | | |   |   |  | |    |||   | ||  | |  |  || ||| || |||
Db  45849 CCCGACCCGGCGCATTCCTCCATGATGGACATCCGGCTGGTCCGCTGCCAGGCTGGTGCC 45790

Qy   1317 GGCGATCCGCTCCCAGTAGCCGGGGGGCGGCACGATGACGCCTGCCGCGCCGAGGACGGG 1376
          | |    |||  |        || | |||         | |||||| || |
Db  45789 GCCCTGACGCGGCTGCACCGAGGAGAGCGATGACGACCCCCCTGCCCCGTC--------- 45739

Qy   1377 TTCGAAGACCAGGGCCGAGACGTTGGGCTTCTCCGCGATGTGCCGGCGCACGAGGGTCGC 1436
                || |   ||||    ||| | ||||||||  | |        |  |    |  |
Db  45738 -----CGATCGACGCCGTCCTGTTCGACTTCTCCGGCACGCTGTTCGCCGTGGAACTGCC 45684

Qy   1437 GCACCGCACGTCGCACGAGGGGTACTCCAGGCCCAGGGGACAGCGGTAGCCAGTAGGGGC 1496
          ||| ||   |  ||  | || |    |||    | | |  ||||   | || |    |
Db  45683 GCAGCGTTGGCTGCGGGCGGCGGCGGCCACCGTCGGCGTCCAGCTCGACCCGGCGCAGCA 45624

Qy   1497 TGTAGCCAGCACGCTGTTGCCGCTGAAGGCCTGGTGGCCGATGTCCCAGTGGACCAGCAT 1556
           |  |||   ||   ||||| ||||  |||| |  | ||   | |     ||| | |||
Db  45623 GGAGGCCCTGACCGGGTTGCTGCTGGCGGCCGGCCGTCC--CGGCGGGCCGGAGCCGCAC 45566

Qy   1557 CCGGGCGCCCATGGTCTTGCCGTGGAAGCCGTGGCGCAGGGCGCAGATCCGGTTGCGGCC 1616
           | | | | |  || || |||| ||    ||     | || |      || |   ||  |
Db  45565 GCCGTCCCGCCGGGGCTCGCCGAGGTGTACGCCCGCCGGGACCTCTCGCCCGAAACGCAC 45506

Qy   1617 CGGCGCGGCGGTCGCCTGGACGACCCGCAGGGCGGCCTCGACCACCTCCGCGCCGGTGGA 1676
          ||  |||           |  | || ||    |  | ||||||  | ||   ||||
Db  45505 CGCAGCG------CGTACGTGGGCCTGCTCTCCACCGTCGACCTGCCCCATCCCGGCCTC 45452

Qy   1677 GAAGAAGGCGTAGGTGTCGAGCTGTTCGGGCAGCAGCCTGGCGAGCAGTTCCAGCAGGCC 1736
          |  || |   |    | |  ||||||| |   | |     |   ||  | ||   | |||
Db  45451 GCCGAGGCGCTCTACGACCGGCTGTTCCGCACGGACGGCTGGCTGCCCTACCGCGACGCC 45392

Qy   1737 GGCGCGGTCCGGCGTGGCGCTGTCGTGGACGTTCCACAGGCGGCGGGCCTGGGTGGTGAG 1796
          | | ||||||    |||||  || | |   |  |  | |    |||  |||    |
Db  45391 GCCCCGGTCC----TGGCGGAGTTGCGCCGGCGCGGCGTGGGCACCGCCGTGGTCAGCAA 45336

Qy   1797 TGCCTCGACGACCTCCGGGTGCCCGTGGCCCAGTGACTGGGTGAGGGTCCCGGCCGCGAA 1856
             | |   |  |    ||  | | |   |     || || ||  ||  | ||  |
Db  45335 CATCGCCTGGGACCTGCGGCCCACCTTCGCGGCGTACGGGATGGCGGAGCTGGTGGACGC 45276

Qy   1857 GTCGAGGTACTGGTTGCCGTCCAGGTCGGTCAGAACGGGACCGCGTCCCTCGGCGAAGAC 1916
           |     | ||       |    || ||||||   | |  || ||   ||     |   |
Db  45275 CTACGTCTTCTCCCACGAGGTGGGGGCGGTCAAGCCCGATCCCCGGATCTTCCGCACCGC 45216

Qy   1917 CCGGCGTCCGTGGACGGCTTCCTCGGAGGCGCCCGGCGCCAGGTGGCGGGCCTCCCGTGC 1976
          | |      |  || ||   || |||| |||  ||  |   |  ||||   | |  |
Db  45215 CTGCGACGAGCTGAAGGTCGCCCCGGA-GCGGGCGCTGATGGTGGGCGACACCCTGGCCG 45157

Qy   1977 CAGGTGCTGTGTCTGCCGTAAGCCTGTCATCGCTGCCTCTGCTCGTCGGACCGGCTGACG 2036
            ||||  | |     ||   |||||    | ||||  ||||  | | | ||| | |||
Db  45156 ACGGTGGGGCGGTGTACGCCGGCCTGCAGACCCTGCTGCTGCCGGCCAGCCCGCCGGACA 45097

Qy   2037 CGATCGCCGGCGAACTGCGTTGTGGCGCACCACGGTTGGGGCGGCTCGGCGCTGAGTCAA 2096
          |  |   |||    |   |  |  | || | |       ||||||  | ||   || |
Db  45096 CCGTGCACGG----CCTGGCCGCCGTGCTCGAACTGGTCGGCGGCGGGCCGGCCAGGCCC 45041

Qy   2097 ACACTTGAACACACACCGCTGCAAGAGTTTGCGGGTTGTTTCAGAAAGTTGTTGC--GAG 2154
           |  |||   | |   ||     | ||    |      |  ||    | ||   |   ||
Db  45040 CCGTTTGCGGAAATGGCGGACGCATAGAGATCCACCGATCACACCTCGGTGGCCCATAAG 44981

Qy   2155 CGGCCCCGGCACTCTGGTTGAGTCGACGTGCTTACGGCGCCACCACGCCTCACGTTCGAG 2214
            |      |||    | |   |||    | | |||  |   |||  || |     |
Db  44980 TAGGGAAATCACGAGTGATCCTTCGGTACGTTCACGAAGAGCCCAGTCCACCGTGACCCC 44921

Qy   2215 GAGGGACCTGTGAGAACAAGCCCGCAGACCGACCCGCTCCCGCGGAGGCCG-AGGTGAAG 2273
            | |  ||     | ||    | |    ||  ||  |||    |||| ||   ||||||
Db  44920 TGGAGTGCTCGTGAACCAGCGACCCGTCTCGGTCCCATCCTTTCGAGGGCGTTCGTGAAG 44861

Qy   2274 GCCCTGGTCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACACCGCCGCC 2333
          || ||||||||||| || ||    ||    ||| || | ||| ||||||||||  |||||
Db  44860 GCGCTGGTCCTGGCGGGAGGGGGTGGATCGAGATTGCGTCCGATCACCCACACGTCCGCC 44801

Qy   2334 AAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTACGCGCTGGAGTCCCTCGCC 2393
          || || ||  | ||| | ||||||||||| ||||| ||||||| ||| ||| |  |
Db  44800 AAACAACTCATTCCCGTGGCCAACAAGCCGGTGCTGTTCTACGGGCTCGAGGCTATTCGT 44741

Qy   2394 GCGGCGGGTGTCCGGGAGGCCGGCGTCGTCGTGGGCGCGTACGGCCGGGAGATCCGCGAA 2453
          | ||  ||| ||  |||||  ||| ||||||| |||     ||  |  || || |
Db  44740 GAGGGCGGTATCGAGGAGGTGGGCATCGTCGTCGGCAGCACCGCGCCCGAAATTCAGCGC 44681

Qy   2454 CTCACCGGCGACGGCACCGCGTTCGGGTTACGCATCACCTACCTCCACCAGCCCCGCCCG 2513
           |   |||||| ||  ||  ||||||  | |   | |||||||| |  |||      ||
Db  44680 GTGGTCGGCGATGGTTCCCAGTTCGGCCTCCAGGTGACCTACCTTCCGCAGGAGGCTCCC 44621

Qy   2514 CTCGGTCTCGCGCACGCGGTGCGCATCGCCCGCGGCTTCCTGGGCGACGACGACTTCCTG 2573
          |  || || |||||||| |||| |||| |||||| || ||| ||||||||||||||| |
Db  44620 CGGGGGCTGGCGCACGCCGTGCTCATCTCCCGCGACTACCTCGGCGACGACGACTTCGTC 44561

Qy   2574 CTGTACCTGGGGGACAACTACCTGCCCCAGGGCGTCACCGACTTCGCCCGCCAATCGGCC 2633
           ||||||| || ||||||| | |   |   ||  |||  ||| ||| |   |  |
Db  44560 ATGTACCTCGGCGACAACTTCATCGTCGGCGGGATCAAGGACATCGTCGAGCGCTTCCGG 44501

Qy   2634 GCCGATCCCGCGGCGGCCCGGCTGCTGCTCACCCCGGTCGCGGACCCGTCCGCCTTCGGC 2693
             || |   ||   || | | |  |||| |||| |||||| ||||||  ||||||||||
Db  44500 CAGGAGCGACCGCACGCGCAGATCATGCTGACCCGGGTCGCCGACCCGCACGCCTTCGGC 44441

Qy   2694 GTCGCGGAGGTCGACGCGGACGGGAACGTGCTGCGCTTGGAGGAGAAACCCGACGTCCCG 2753
          ||||| ||| | | |||||||||    ||| |  |  |||||||||| |||||   |||
Db  44440 GTCGCCGAGATGGGCGCGGACGGCCGGGTGATCGGGGTGGAGGAGAAGCCCGAGCACCCC 44381

Qy   2754 CGCAGCTCGCTCGCGCTCATCGGCGTGTACGCCTTCAGCCCGGCCGTCCACGAGGCGGTA 2813
             |||   || || ||  |||| || |||| |||||||||||  |||||||| || ||
Db  44380 AAGAGCGACCTGGCCCTGGTCGGGGTCTACGTCTTCAGCCCGGTGGTCCACGATGCCGTG 44321

Qy   2814 CGGGCCATCACCCCCTCCGCCCGCGGCGAGCTGGAGATCACCCACGCCGTGCAGTGGATG 2873
             |   | |  || |||   |||  |||||||||||||||| |||| ||||||||||||
Db  44320 GCCGAGCTGAAGCCGTCCTGGCGCAACGAGCTGGAGATCACCGACGCGGTGCAGTGGATG 44261

Qy   2874 ATCGACCGGGGCCTGCGCGTACGGGCCGAGACCACCACCCGGCCCTGGCGCGACACCGGC 2933
          |||||||| ||||  || ||     ||   |  | |||  |   ||||   |||||||||
Db  44260 ATCGACCGCGGCCACCGGGTCGACTCCACCATGATCACGGGCTACTGGAAGGACACCGGC 44201

Qy   2934 AGCGCGGAGGACATGCTGGAGGTCAACCGTCACGTCCTGGACGGACTGGAGGGCCGCATC 2993
          | |   |  |||||||||||| | |||||   |||||| | | |  | |||       ||
Db  44200 AACCTCGCCGACATGCTGGAGATGAACCGGTTCGTCCTCGGCAGCGTCGAGCCTTTGGTC 44141

Qy   2994 GAGGGGAAGGTCGACGCGCACAGCACGCTGGTCGGCCGGGTCCGGGTGGCCGAAGGCGCG 3053
          ||||| ||||||||||  | || |   ||| |||||| |||   | |   ||  |||||
Db  44140 GAGGGCAAGGTCGACGACCGCACCGAACTGATCGGCCCGGTGGTGATCAGCGCCGGCGCC 44081

Qy   3054 ATCGTGCGGGGGTCACACGTGGTGGGCCCGGTGGTGATCGGCGCGGGTGCCGTCGTCAGC 3113
            |||  | || || |   | || |||||||||||| |||| || ||  |  || || ||
Db  44080 GACGTCAGCGGCTCCCGGATCGTCGGCCCGGTGGTGGTCGGAGCCGGCTCGATCATCCGC 44021

Qy   3114 AACTCCAGTGTCGGCCCGTACACCTCCATCGGGGAGGACTGCCGGGTCGAGGACAGCGCC 3173
          ||||||    |||||||||||||||| ||||     ||||||    |   ||| | ||
Db  44020 AACTCCTACCTCGGCCCGTACACCTCGATCGACTGCGACTGCACCCTGCTGGAGACCGAG 43961

Qy   3174 ATCGAGTACTCCGTCCTGCTGCGCGGCGCCCAGGTCGAGGGGGCGTCCCGCATCGAGGCG 3233
          |||||| ||||| || |||||||||| ||| |  |||||||      ||||||||||
Db  43960 ATCGAGCACTCCATCGTGCTGCGCGGTGCCTACATCGAGGGCATCGGCCGCATCGAGTTC 43901

Qy   3234 TCCCTCATCGGCCGCGGCGCCGTCGTCGGCCCGGCCCCCCGTCTCCCGCAGGCTCACCGA 3293
          ||  | |||||||| |  ||    || | ||| |  |||||  | ||  || | |||||
Db  43900 TCGATGATCGGCCGGGAGGCGCGGGTGGTCCCCGGGCCCCGGGTGCCCAAGACGCACCGG 43841

Qy   3294 CTGGTGATCGGCGACCACAGCAAGGTGTATCTCACCCCATGACCACGACCATCCTCGTCA 3353
           | ||  ||||||| |||||| | ||     |       |||       | |||| ||||
Db  43840 TTCGTCCTCGGCGATCACAGCGAAGTCCGGGTGGGGGTCTGATGCGTCGCGTCCTGGTCA 43781

Qy   3354 CCGGCGGAGCGGGCTTCATTCGCTCCGCCTACGTCCGCCGGCTCCTGTCGCCCGGGGCCC 3413
          ||||||| || || |||||  ||||    |||||||||   ||     |    |    ||
Db  43780 CCGGCGGTGCCGGTTTCATCGGCTCGCAGTACGTCCGCGACCTGGCCACCGGTGCCTACC 43721
```

```
KW   gene cluster; genetic manipulation; contig; gene; ds.
XX
OS   Micromonospora carbonacea.
XX
PN   WO20015180-A2.
XX
PD   02-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-CA00128.
XX
PR   27-JAN-2000; 2000US-0177711.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
PA   (FARN/) FARNET C.
XX
PI   Staffa A, Zazopoulos E, Mercure S, Nowacki P;
XX
DR   WPI; 2001-476185/51.
DR   P-PSDB; ABB06898, ABB06899, ABB06900, ABB06901, ABB06902, ABB06903,
DR   ABB06904, ABB06905, ABB06906, ABB06907, ABB06908, ABB06909, ABB06910.
XX
PT   Novel isolated gene cluster encoding polypeptides involved in
PT   everinomicin biosynthesis useful for construction of everinomicin
PT   overproducing strains, and to allow chemical modifications of
PT   everinomicin to enhance certain properties -
XX
PS   Claim 1; Fig 1; 181pp; English.
XX
CC   ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora
CC   carbonacea everinomicin biosynthetic locus gene cluster. The contigs
CC   encode the protein sequences designated ORF (open reading frame) 1 to
CC   49, given in ABB06881 to ABB06930. The gene cluster is useful for the
CC   construction of the everinomicin antibiotic in overproducing strains,
CC   and to allow chemical modifications of everinomicin to enhance certain
CC   properties via genetic manipulation or combinational biosynthesis. The
CC   gene cluster can be used to produce genetic systems and genes encoding
CC   novel enzyme activities, and avoid the problems of low yield and quality
CC   of everinomicins produced by chemical synthesis.
XX
SQ   Sequence 14186 BP; 2084 A; 4892 C; 5016 G; 2194 T; 0 other;

Query Match             11.9%;    Score 817.2; DB 23;    Length 14186;
Best Local Similarity   63.7%;    Pred. No. 8.1e-93;
Matches 1296;   Conservative    0;  Mismatches 723;   Indels   15;   Gaps     3;

Qy   2268 GTGAAGGCCCTGCTTCCTGGCCAGGTGGAACCGGCCAGCAGCAGACTGAGGCCGTTCACCCACACC 2327
          ||||| ||||  | || | ||||| |||| |||| |||||||||||| || | ||||||| |||||
Db  10916 GTGAAAGCGCTGGTATTGGCGGGTGGAATCGGCTCGCGAATGCGCCCGATCACCCACACG 10975

Qy   2328 GCCGCCAAGCAGCTGCTCCCCATCGCCAACAGCCCCGTGCTCTTCTACGGCGTCGGAGTCC 2387
          | || | | |||||||| |  ||| ||||| |||||| || |||||| |||||  |||||| 
Db  10976 TCAGCGAAGCAGCAGCTCATTCCGGTCGCGAACAAAACCGGTCCTCTTCTACGGCCTGGAAGCA 11035

Qy   2388 CTCGCCGCGGCGGCGTGTCCGGAGGCCGGCGTCGTCGGCCACCGCGGCCGGAGATC 2447
          |||| |  | || | | | |||| | | |||| || ||| | ||||||| || ||
Db  11036 ATTCGTGACGCCGGGCTCGGCGCTTGAAGGTGACCTACCTGGCGCATCGCGGCGGAGATC 11095

Qy   2448 CGGGAACTCACCGGCGACGGACCCCCGGCGTTCGGGTTACGGCACTCACCTACCTCCACCAGCCC 2507
          |||||| |||||  ||||||| |||||||||| ||||| ||| ||| ||||| ||| |||| |||
Db  11096 GAGCGGGGCGGTCGGTGACGGCTCGCAGTTCGGCTTGAAGGTGACCTACCTGGCGCATCGCGGCGG 11155

Qy   2508 CGCCGCTCGGGTCTCGGCCGGTCTCGGTGGGCAGCGGTGCGCATCGCCCCGGGCATCGGCGG 2567
          |||||||| ||  ||||||| |||||| || | || |  |||| | |||| ||| |  ||||
Db  11156 GCCCGGGCGGTCTCGGACGCCGGTGGGGGGCCGGCGACAGCCGCGTCACCGGGTCCCGGACGGAC 11215

Qy   2568 TTCCTGCTGTACCTGGGGACAACTACCTGCCCGGCGACGACTTCACCGGCTTCCGCCCGCCAA 2627
          ||  || ||  |||||| |||   | | |||| |||||| |  || ||| || | ||| || |
Db  11216 TTCCTGATGTACCTGGGCACAACTTCGTCCTCGTGGTGCATCAACGACGCGGTCGAGCGG 11275

Qy   2628 TCGGCCGCGAGCCCGATCCCGTGGCGGGCCGCCGGCGGCGTCACCCCGGTCGCGGACCGTCCCCC 2687
          |||||||| ||   |  ||| | ||| | |  |||| | |||| ||||| |||| ||| |||| 
Db  11276 TTCCGCCGGAACGCCCGCACCCGCGCGGCCCAGCTGATGCTGACCGATGCTGACCAAGGTCAAGGATCCGCACGCC 11335
```

Qy   3414 CCGGC------GGGCGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCCGGCAGCCTCG 3467
          |||        | |||||| |||| | | || | |||||||||| ||||  || ||||||
Db  43720 CCGACAACGGCGGCCGGCGCAGGCGCGGTGACCGTCGCGGCGGCAACCTCG 43661

Qy   3468 CCCGGCCTGCACGCGGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGCGACGTGTGCCG 3527
          |  || | ||  |||| ||| ||||| |  || | |||| | ||||| |||| ||||||||
Db  43660 CCAACCTCGAACCGGTCCAGGACC------GGATCACCTTCGTCCAGGGCGACGTCTGCGG 43607

Qy   3528 ACACCGCGCTCGTCGACACGGCTGGCCGCCGCGGCACAGCATCGTGCACTTCCGCGCCG 3587
          | ||| | | |  | || | ||  |    | | |||| | |  ||||| |||||| |||
Db  43606 ACACGGCGCTGCTGGCCGAGTGCTGCGCCACGACGTGGTCAACTTCGCCGCCGCC 43547

Qy   3588 AGTCGCACGTCGACCGCTCCATCACCGACAGCGGTGCCTTCACCCGCACCACACGTGCTGG 3647
          ||||||||||||||  || |||||||| ||| | | |||||| ||||| | ||||| |||
Db  43546 AGTCGCACGTCGACCGGTCCATCGCCGACTTCGTCGGCGAGTTCGTCCCGCACCAACGTGCAGG 43487

Qy   3648 GCACCCAGGTCCTGCTCGACGCCGGCTCCGCGCACCGTGCGCGCACCTTCGTCGCACGTCT 3707
          || ||| | | | | ||  |  | | |  |  | ||||||  || ||| |||  |||| ||
Db  43486 GCGTCCAGACGCTCATGCAAGCCGGCGTGTCTGCACCCGGGTCGGCCGGGTTGGTCCAGGTCT 43427

Qy   3708 CCACCGACGAGTGTACGGCGTCTCCTCCCGCACGGGGCCGCGGGAGAGCCGACCCCCTGC 3767
          || |||||||| | || ||| | || |  |  | | || | | |||||| || || |||
Db  43426 CCACCGACGAGTCTACGGCGTCACGAGCAGCATCGACGAGGGTTCCTGGGCGGAGCACCCCCTG 43367

Qy   3768 TTCCGACCTCGCCGTACGCGGCGTCGAAGGCGCGCTCGAAGGCGGACTCATGGCGCTCGCCACC 3827
          |||| |  | ||| || ||| ||||| ||  | |  ||||| |  | |  | |||| |  |
Db  43366 CGCGGAACTCCCCGTACGCGGCGGCGGCGAAGGCCCGGCCGGCGACCTGATCGCCCTGGCCTACG 43307

Qy   3828 ACCGCACCCCACCGGCGGCTCGACGTCCGGGTGACCCGGTCGGCCCCCACC 3887
          | |||||||| | || | ||  || || | |  ||  | | |||| || |
Db  43306 CCCGGACCCCACGGGCTGCCGGTCTGCCTCACCCGCGCCGGCCAACAACTACGGGCCGTACC 43247

Qy   3888 AGCATCCCGAGAAGCTCATACCCGGCTTCATACCCGTCCGGCGGCGCCACCGTTC 3947
          || |  |||||||||| | ||  |  || || | |   | | | |  |  || 
Db  43246 AGTTTCCGGAGAAGCTGATCCCCGTGTCGTCACCGAGCTGTTGAACGGGCGACGGGTGC 43187

Qy   3948 CCCTCTACGGCGACAGCGGCGACGTGCGCGACTGGCTGCAGTCGACGTCGACGCACGTCAGGG 4007
          |||  |  || | || | | | | || ||   | | |||| || |||| |||| | |||||
Db  43186 CGCTGTACGGCGACGGCGGCGCGGGAACCTCCGGCGACTCGCGACTGGCTGCAGTCCACGTGACCGGCGCCGGG 43127

Qy   4008 CCGTCGAACTCGTCGCCGGCGTGTCGGCCGCCGGCCGGCGGAGAGATCTACACACATCGGGGGCGGCGCA 4067
          || | ||  |||| | |  |  || | | | | | | |||||| |  | ||||||  |
Db  43126 GCATCCAGACCGTCGTCGACCCGGTGTCCCGGCGTGTCCGGAGGTCTACACCATCGCCGGCGCACGCACGG 43067

Qy   4068 CCTCGCTGCCCCAACCTGAGCTCACGCACCGGCTCGCACTGTGCGCCGGGCGGCCCGG 4127
          ||| | | ||  || | |||| |  | ||| |  |  || | || || ||| ||||||
Db  43066 CTGAGCTGTCCAACATGGAGCTGCACCGGCACCGGCTCGCACTGTGCGCGCCTGGGGGGCCGGGTGGG 43007

Qy   4128 AGGCCATCGTCCACGTCGAGAACCGCAAGGCGCACGACCCGCTACGGCGGTCGACCACA 4187
          | ||||||| | | | | | ||  | ||| ||| | || | | |  || | || || |
Db  43006 ACCGGGTCGAGCGGGTGCTGCGACCGGCAAGAGGGGCCCAAGGGCCGCGACCTACTCGCTGACGACG 42947

Qy   4188 GCAAGATCACCGGCGAACTCGGTTACCGCGGTTACCCGGACTTCGGGACGCTGCGCCG 4247
          || | | |||| ||  | | | | |||||||| |  ||| |  || | ||  || | |
Db  42946 CGAAACTGCGGGCG---CTCGGCTACCGGGCG--CTCGGCTACCGCCTTCGCCGTCACGCGGCCTGCGCCG 42890

Qy   4248 ACACCGCGAAGTGCTACGAGCGGCACAGGACTGGTGCGTCCCCTG 4294
          | ||| | | |  | |||  |  | |  |||| ||||| | || |
Db  42889 AGACGATCGACTGGGTACGGCACGCACCGGCACCGGGACTGTGGGAGCCGCGGCTG 42843

```
RESULT 3
ABL50559
ID   ABL50559 standard; DNA; 14186 BP.
XX
AC   ABL50559;
XX
DT   18-JUN-2002 (first entry)
XX
DE   Micromonospora carbonacea everinomicin locus nucleotide contig 5.
XX
KW   Micromonospora carbonacea; antibiotic; everinomicin; biosynthesis;
```

Right column (continuation):

```
Db  12410 CGCTGGCGCCGAACTCGCCGCGGCGCCAAGCGGCGGGGTGACCTGATCGCCCTGG 12469
Qy   3822 CCCACCACCGCACCCACGGCCTGGACGTCCGGGTGACCCGCTGGTTCGAACAACTTCGGCGC 3881
Db  12470 CGTACGGCGGACGTACGGACGTCCGGTCCATCACCAGTGCGCAACACTACGGTC 12529
Qy   3882 CCCACCAGCATCCCGGGAAGCTCATACGCCGCTTCCTGACCAGCCTCCTGTTCCGGCGGCA 3941
Db  12530 CATACCAGTTCCCGGAGAAGGTGATCCCCTTCCTCCACCCGTCTGATGGACGGCGGT 12589
Qy   3942 CCGTTCCCCTCTACGGCGACGGCGCACGTGCGGCGACTGGCTGCTCGACGTCGACGACACG 4001
Db  12590 CGGTCCGCTCTACGGCGACGGCGCCAACGTCCGCGACTGGCCGATCCACGTGGCCGACCACT 12649
Qy   4002 TCAGGGCCGTCGAACTCGTCCGCGTGTCGGGCCGCCGGCGGAGAGATCTACAACATCGGGG 4061
Db  12650 GCCGTGGCATCCAGACGTGCTGCGACGCGGTGCGTCCGCGAGGTCTACCACCATCGCCG 12709
Qy   4062 GCGGCACCTGCTGCCAACCTGGAGCTCACGCGGCGTTGCTCCACTGTGCGGCGCGG 4121
Db  12710 GGACGGCCGAGCTGACCAACCTGGAACTCACCCAGCACCTGCTGACACGCCGCGGTCGGCGGAA 12769
Qy   4122 GCCCGGAGCGGCGCATCGTCCAGTCGAGAACCGCAAGGGGCACGACCGCGCTACGCGGTCG 4181
Db  12770 GCTGGGACGGCCGTCGAGAGGGTGCGCACCGCTAAGGGCCACGACCGCGTACTCGCTTT 12829
Qy   4182 ACCACAGCAAGATCACCCGCGGAACTCGGTTACCGGCGCCACCGACTTCGCGACCGCGCGC 4241
Db  12830 CCGACGCGAAGCT---CCGGGCCCTGGGCTACGCCCCGCGTCCCCTTCGCCGACGGCC 12886
Qy   4242 TGGCCGACACCCGGAGTGGTACGAGCGGCACGAGGACTGGTCCACTGTGGAGCCGCTGC 4295
Db  12887 TGGCCGACGGTCGCGTGGTACCCGCGGAACCGGCACTGGTGGAGCCGCTGC 12940
```

```
RESULT 4
ABA99469/c
ID   ABA99469 standard; DNA; 38064 BP.
XX
AC   ABA99469;
XX
DT   31-MAY-2002 (first entry)
XX
DE   Actinoplanes sp SE 50/110 (CBS614.71) DNA encoding acarbose operon.
XX
KW   Acarbose; AcbAAsp; AcbBAsp; AcbCAsp; AcbDAsp; AcbEAsp; AcbFAsp; AcbGAsp;
KW   AcbHAsp; AcbIAsp; AcbJAsp; AcbKAsp; AcbLAsp; AcbMAsp; AcbNAsp; AcbOAsp;
KW   AcbPAsp; AcbQAsp; AcbRAsp; AcbSAsp; AcbTAsp; AcbUAsp; AcbVAsp; AcbWAsp;
KW   AcbXAsp; AcbYAsp; AcbZAsp; Asp3-1; Asp3-2; Asp3-3; bioconversion;
KW   biosynthesis; alpha-glucosidase inhibitor; ds;
XX
OS   Actinoplanes sp.
XX
PN   DE10021667-A1.
XX
PD   08-NOV-2001.
XX
PF   05-MAY-2000; 2000DE-1021667.
XX
PR   05-MAY-2000; 2000DE-1021667.
XX
PA   (FARB ) BAYER AG.
XX
PI   Apeler H, Wehlmann H, Piepersberg W, Diaz-Guardamino P, Jarling M;
PI   Thomas H, Wehmeier U;
XX
DR   WPI; 2002-227993/29.
XX
PT   New nucleic acid sequences from Actinoplanes, useful for synthesis and
PT   bioconversion of acarbose and related inhibitors of alpha-glucosidase
PT   -
XX
```

Left column (main alignment):

```
Qy   2688 TTCGGCGGTCGCGGAGGTCGACGCGGACGGGAACGCTGTGCGCTTGGAGGAGAAACCCGAC 2747
Db  11336 TTCGGCATCGGCGACGATCGGGCCCGGACGCCGGGTCGTCGATGTCGAGGAGAGCCCCG 11395
Qy   2748 GTCCCGCCGCAGTCGTCGCTGCATCGCGTCGGGGTGTTCAGCCGTTCAGCCTCCGGTCTCCACGAG 2807
Db  11396 TATCCCAAGAGCGACCTCGCTCGTGGTGGCGTGTACGTGTTCAGCCCGTCGTGCACGAG 11455
Qy   2808 GCGGTACGGGCCATCACCCCCTCCGCCCGGCGAGCTGGAGATCACCACCGCCGTGCAG 2867
Db  11456 GCGATAGCCGGAACTGAAGCCGTCGTGGCCCAACGAACTGGAGATCACCGACGCCATCCAG 11515
Qy   2868 TGGATGATCGACGGGGGCTGCGCGTACGGGCGCGAGACCACCACCCGGCCCTGGCGCGAC 2927
Db  11516 TGGCTGATCGACCACGACAGGCGCTATCGAATCACCATAATCACCGGATTCTGAAGGAC 11575
Qy   2928 ACCGCAGCGCGCCGGAGGACGATGGGGAGGTCAACACCGTCACGTCCTCGACGGACTGGAGGGC 2987
Db  11576 ACCGGCAGCCCTGCCGACATGCTGGAGATGAACCGGTTCATCCTGGAAAGCCTCGACTCC 11635
Qy   2988 CGCATCGAGGGGAAGGTCGACGCGCACACCACGTCGGTCGGTCGGGGTGGCCGAA 3047
Db  11636 GAGGTGAGTGGCGAGTCAGTGCGACACCGAGATCACCGGTCGGGTCGTGATCGGGCCC 11695
Qy   3048 GGCGGATCGTCGCGGGGTGCACACGTGTTGGGCCCGGTTGGTGATCGGCCGGGTGCCGTC 3107
Db  11696 GGGGCGGTCATCACCGGGTCGCGGGATCATCGGGCACGCCCGTCGTGCGGGGTCGCGATC 11755
Qy   3108 GTCAGCAACTCCAGTGTCGGCCCGTACACCTCCATCGGGGAGGACTGCCGGGTCGAGGAC 3167
Db  11756 ATTCGCAACTCGCAGCTCGGCCCGTTCACGTCGATCGACTGCACCGTCATCGAC 11815
Qy   3168 AGCGCCATCGAGTACTCGTCCGCTGCTGCGGCGCCCAGTCGAGGGCGTCCCGCATC 3227
Db  11816 AGCGAGATCGAGCAGTCCATCGTCGCGTTCATGCCGCCTTCATCGACGACGATCGGCCGGATC 11875
Qy   3228 GAGGCGTCCCTCATCGGCCGCGCGCGTCGTCGGCCGGGTCTCTCCCGCAGGCT 3287
Db  11876 GAGTGGTCGATGATCGGGCCGTGAGGGCGCGCGCTGACCCCGGGCGCGCCAGAACG 11935
Qy   3288 CACCGACTGGTCGGCGACACCAGCAAGGTGTATCTCACCCATGACCACGACCATCC 3347
Db  11936 TACCGCTTCGTTCGTCCTGGCGACCACAGTGAAGTACGGGTTAGGCGTGTAGTGCCGAGGGTCT 11995
Qy   3348 TCGTCACCGGCGGAGCGGGCTTCATTCGGCCTCCGGCGCTTCCGGCTCTCGCCCG 3407
Db  11996 TCGTGGCCGGTGGCGGCCGGCTTCATCGGGCTTCATCGGGTCTCGCGCGGAACTCGTCGCCGCCGGGG 12055
Qy   3408 GGGCCCCCGGC------GGCGTCGCGGTGACCGTCCTGACAAACTCACCTACGGCCGGCA 3461
Db  12056 CGTACGGCCGGGTCGGCGTGGCGAGGGCTGCGCAGGTCACGCGTCACGTCACCTATCGGGAA 12115
Qy   3462 GCCTCGCCGCCGCGCCCTGCACGCGGTGCGTGACCATCCGGCCTCACCTTCGGCAGGGCGACG 3521
Db  12116 ACCTGCGGCGGAATCGCCGCGGGTGCGGGA-----CGCCGTCACCTTCGTCCGCCGGGTGACA 12169
Qy   3522 TGTGCCGACACCGGTCGTGCGACCAGCGAAGGTGCTGGCCGCCGGACGACATCGTGCACTTCG 3581
Db  12170 TCTGCGACGGCCACGACTGCTCCCGGCGGCCACCGACGAGCGTGTGCTGAACTTCG 12229
Qy   3582 CGGCCGAGTCGCACGTCGACCCCGTCCATCGAACAGCCGGTGCCTTCACCCGCACCAACG 3641
Db  12230 CGGCCGAGACCCACGTGTCGACCGTCCATCGCCGGACTTCCGCGGACCAACG 12289
Qy   3642 TGCTGGGCACCCAGGTCGCTGCGACCGGTGCGTGTGCCCACGGTGTGCCGCCACCTTCGTGC 3701
Db  12290 TTCAGGGCGGTCCAGTCGCTACGGCCAGCCGGTCTACCGGCCGATCCGGAGCGGAGCGCGGC 12349
Qy   3702 ACGTCTCCACCGACGAGGTGTACGGCGTACGGCGTCCTCCCTCCCCGCACGGACC 3761
Db  12350 AGGTCTCCACCGACGAGGTGTACGGCGATCCTGGACGCGGAGGACGCGGC 12409
Qy   3762 CCCTGCTTCCGACCTCCGCCGTACGCGGCGCCTCGAAGGCGGCCTCATGGCGCGTCG 3821
```

PS      Claim 1; Page 67-77; 80pp; German.
XX
CC      This invention describes novel nucleic acid sequences from
CC      Actinoplanes sp. SE50/110 (CBS 614.71) designated acbA, B, C, D, E, F,
CC      G, H, I, J, K, L, M, N, O, P, Q, R, S, U, V, W, X, Y and Z, and asp3.1,
CC      3.2 and 3.3, and their homologues. The products of the invention
CC      individually or collectively, are used for synthesis or bioconversion of
CC      acarbose (or its precursors or related substances with alpha-glucosidase
CC      inhibiting activity), especially of alpha-glucosidase inhibitors. The
CC      products can also be used for optimising/inducing production of such
CC      compounds in Actinoplanes or other organisms. This sequence represents a
CC      DNA fragment which encodes the genes involved in the acarbose
CC      biosynthesis and bioconversion pathway which are described in the
CC      invention.
XX
SQ      Sequence 38064 BP; 5622 A; 13333 C; 13603 G; 5506 T; 0 other;

Query Match             10.9%;  Score 745;  DB 24;  Length 38064;
Best Local Similarity   72.4%;  Pred. No. 5.4e-84;
Matches  980;  Conservative    0;  Mismatches  370;  Indels    3;  Gaps    1;

Qy      739 GCCCTCTCCAGCAGGCGCCAGGCCCTCGTCGAGCAGCGCCTCTGTCGAGGGTGAACGGCGGT 798
                |||  ||| |||||||  ||||||            ||  | |||||||| ||||||||
Db     5878 GCCCGGTCCAGCAGGTCCAGGCCGCGCCAGCTCCGTCTCGTCGATGACGAACGGCGGC 5819

Qy      799 GCCAGCCCAGGATGTGGCCGCGCCCAGGGAGGTGCGCCAGCCCAGGTCGAGGGCGGTGGTG 858
                || ||| ||||| |||||  |||| | |||||||||||||||||||||||||||||
Db     5818 GCGAGCCCGGATGATGTGCCCCGCGATGCGCGTGGCGGTGCGGAGGCGCGGTGGTG 5759

Qy      859 TAGACGGCCGCGGGCGGTCTCGGGGCGGTGCGGTGCCCGGCGTCGGGTGGTGACGAACTCC 918
                  ||| ||||| | ||||||  | ||  |||| |||| | |||||||| |||||
Db     5758 TAGACCCGGCCGGGCGATCTCCGGGGCCGCGGCTCCGGCGTCGGGTGCGGTGCACGAACTCG 5699

Qy      919 AGGCCCCACAGCAGTCCGAGGCCGCCGTACCTGGCCGAGCTGGGGAAGCGGGACTCCAGG 978
                |||  || ||| |  ||||||  | ||| | ||||||||| ||   || |||  | ||
Db     5698 AGCCCGTGCAGCCGCGATGCCCGCCGTCCCCAGTGCGCCACGCGTGCCCGGTGCGTCGCATC 5639

Qy      979 GCGGCCAGGCGGTCTCGGCCAGCACCCGGTCGGCGCACCCGCGAGTCCGGACCCGCGTCG 1038
                | | ||| | |    || ||| ||||  ||||| | | ||||||||||||| ||||||
Db     5638 TCGGCCAGGCGGTCTCGGCCAGCACCCGGTCCGCGGCACCCGGACCCGCGAGTAGCCGCGC 5579

Qy     1039 CGCTCGGACGACCTCCAGCGTGGGCGGCGGGCGGGCGGCGATCCCCAGTGGGTTGCTCGCGTAC 1098
                || |   || |||| || |||| ||||  |||||||  || ||||||| |||
Db     5578 CGGGAGATCACCGAGAGGGGTGGCGTGGGCGGGCGGCGATCCCGAGACGGGCGGGTTAG 5519

Qy     1099 GTCGAGGCGTACGGCCCCGGGCGGGTGCCCGCCTCCGG 1155
                || ||||| ||| |  | ||||| | ||| | | ||||
Db     5518 GTGGAGGCCGGTCGATCGGCGGCGCGGAGCCCGGGATGCGCAGCCCGGTTCGTCACGCCCG 5459

Qy     1156 GCCAGCACGGCGAAGGCGAATCCGCTCGGCGTGCCCTTGGACAGCAGCATCGCCCAGGTCCGGC 1215
                |||||||  | ||   | ||||||||| || |||  || ||| |  |  | |||||| | || ||
Db     5458 GCGAGCACGGCGAACGGCGAAACCGGACGCGGTGCCCTTGGCCAACGTGACCAGATCCCGGG 5399

Qy     1216 TCGATGCCGAACAGTTCGCTGGCGGAGGAAGTTCCGGTCGCCGGTGCTCCGGTCGAGCACC 1275
                |||  |  |||| || |  ||||| |  || | ||||| |||||||
Db     5398 GCGACCCGATCGGCTCGCTGGCGGCGCGCGTTCCGGTCCGGTCCGGCGCCGCGGTGAGCACC 5339

Qy     1276 TCGTCGGCGACGAGCAGCAGCACGCCCGCGGCGGCGGGCGGATCCCGTCCCCAGTAG 1335
                |||||| || || |  ||||| |||| ||||||||| || |||| || | |||||||||
Db     5338 TCGTCGGGCACCAGCAGCACGCCGTTGTCCGCGCGGCGCGACGCGGTTCGAAGACCGGCCAGTAG 5279

Qy     1336 CCGGCGGGGGGCGGCACGGATGGCCGCGATGAGCTGGCCGCGGCGGTTCGAAGACCAGCCGGCCGAG 1395
                |||| || | ||| |  || |||| ||||  ||||| |||||  ||| ||||||
Db     5278 CCCGGTGCCGGCGACGATCACCGCCGGGCGCGGGCGCCACCGCCGCCGCCCAGGCCGCCTC 5219

Qy     1396 ACGTTGGGCTTCTCCGCGATGTGCCCGGCGTCGCACGAGGGTCGCACGAGCTCGGCACGAG 1455
                || |||||| |||| |||||  |||||||||  || |||| || |  | || ||||||| |
Db     5218 ACGTTGGGTTTCTCCGCGATGTGCCGGCGTGCCCGGCGTACCAGGGTGGGCCATCGGACACGACACC 5159

Qy     1456 GGGTACTCCAGGCGCAGGGACAGCGGTAGCCAGTAGGGCCTGTGTTG 1515
                |||||| |||| ||||| |  ||| | ||||||  || ||||||
Db     5158 GGATAGCTCAGCTTCCAGCTCCAGCAGCGTACGGGCGCCGTACCCAGCGTGTTG 5099

Qy     1516 CCGCTGAAGGCCTGGTTGCCGATGTCCCAGTGGACCAGCATCCGGCGGCGCCATGCTCTTG 1575
                ||| ||| || |||||   ||   | |||||||| |||||||| |||||||||||||||||
Db     5098 CCGGCGAACGACTGGTCCCCCAGTGCACCAGCATGCCGGCGCCCCATCCGCTTG 5039

Qy     1576 CCGTGAAGCCGTGGCCAGCGCAGATCCGGTTGCGCGCCGCGGCGGCCCGGCCGCGCCCTGG 1635
                |||||| || |||||| |||||  |||||||| |||||||||||   |||||
Db     5038 CCGTGGAGCCGTGCCGCCGGTTGCGCGCCGGTTCGGCGGACGGCCTGG 4979

Qy     1636 ACGACCCGCAGGGCGCCTGACCACCTCCGCGCCGCGCCGTGGAGAAGGCGTAGGTGTCG 1695
                ||   ||||||  | ||    |  |||  || |||||||||||||  |   |   ||
Db     4978 ACCGTCCCAGTGCCGCTTCCACCACCTCCGAGCGCCCGTGGAGAAGAAGGCGAGCGTGGTG 4919

Qy     1696 AGCTGTTCGGGCAGCAGCCTGCGAGCAGTTCCAGCAGCGGCGGGCGCGGCGTCCGGCGTGGCG 1755
                || ||||||||||| | |  |||||||||||||||||   ||||||||| ||||| |||
Db     4918 AGGTGGTCGGGCAGCAGTTCGGGCAGTTCCAGCAGCGGCGGGCGCGGGCGGTCGGGCGGTGCGCG 4859

Qy     1756 CTGTCGTGGACGTTCCACAGCCGCGGCGTCGGGCGTGGTGGTTGAGTGCCCTCGACGACCTCCGGG 1815
                |||||||||||||||||||  |||  |  |  |||||||| ||| || ||  |||| |  |||||
Db     4858 AAGTCGTGGACGTTCCACAGCTTCCCGGCCTGTTCGGGTGAGGCCGCACCACCTCCGGG 4799

Qy     1816 TGCCCGTGCCCAGTGACTGGGTAGGGTCCCGGCCTGTTCCGGCGCGGCGGCGTATTCCACACCG 1875
                |||| || ||||  | ||   || || || ||||||||||||||||||||||| ||  |
Db     4798 TGGCAGTGCCCAGTCAGTCGACTCAAAGTACCGGCGGCAAGTCGAGGTATTCGAGGTATTCCACACCG 4739

Qy     1876 TCCAGGTCGTCGAAACGGGACCGCTCCCTCCGCGCGAGACCCGGCTCCGTGGACGGCT 1935
                ||  |||||| ||| ||||| ||| |  |  |||| |||||||||||||||||||| |
Db     4738 TCGGGCTCGCGCAGGGTCCGGCGTCGCGGCCCCGGCCCGAGCGCCCCC 4679

Qy     1936 TCCTCGGAGGCGCCCGGCGCCCAGGTGGCGGCGCCTCCCGTGCGTCCGGGGCCGCCTGCCTG 1995
                | |  ||||| ||| |||||| ||||  |||||| | |||||||||||||| |  ||  |
Db     4678 TCTTCGGAGGCTCCCGAGCCAGGTCAGGTGACCGACTCCCGGCGAGCCAGTCGGCGCCGAACTCG 4619

Qy     1996 AAGCCTGTCATCGCTCGCCTCTGCCGGTCGTGCGGCGGGCAGCTCCAGG 2055
                ||||  |  | | ||  | |   |  | || |||||||||||  | ||  ||
Db     4618 CTCACCGGAACCGCCGCCGGTCGGCGCCGGTGTCGAAACGAACCGGCTCGGAACACA 4559

Qy     2056 TTGTGGCCACCACGGTTGGGGCGGGCCGTTCGGCGGC 2088
                ||| ||  ||||| || ||| |||   |  |||||||
Db     4558 TCACGGGTCACTCCTGAAGGGGATAGCCGCCGCGC 4526

RESULT 5
ABA99445/c
ID      ABA99445 standard; DNA; 1350 BP.
XX
AC      ABA99445;
XX
DT      31-MAY-2002 (first entry)
DE      Actinoplanes sp DNA encoding acarbose synthase AcbVAsp.
XX
KW      Acarbose; AcbAAsp; AcbBAsp; AcbCAsp; AcbDAsp; AcbEAsp; AcbFAsp; AcbGAsp;
KW      AcbHAsp; AcbIAsp; AcbJAsp; AcbKAsp; AcbLAsp; AcbMAsp; AcbNAsp; AcbOAsp;
KW      AcbPAsp; AcbQAsp; AcbRAsp; AcbSAsp; AcbUAsp; AcbVAsp; AcbWAsp;
KW      AcbXAsp; AcbYAsp; AcbZAsp; Asp3-1; Asp3-2; Asp3-3; bioconversion;
KW      biosynthesis; alpha-glucosidase inhibitor; ds.
XX
OS      Actinoplanes sp.
XX
PN      DE10021667-A1.
XX
PD      08-NOV-2001.
XX
PF      05-MAY-2000; 2000DE-1021667.
XX
PR      05-MAY-2000; 2000DE-1021667.
XX
PA      (FARB ) BAYER AG.
XX
PI      Apeler H,  Wehlmann H,  Piepersberg W,  Diaz-Guardamino P,  Jarling M;
PI      Thomas H,  Wehmeier U;
XX
DR      WPI; 2002-227993/29.

XX
PT   New nucleic acid sequences from Actinoplanes, useful for synthesis and
PT   bioconversion of acarbose and related inhibitors of alpha-glucosidase
PT   .
XX
PS   Claim 1; Page 10; 80pp; German.
XX
CC   This invention describes novel nucleic acid sequences from
CC   Actinoplanes sp. SE50/110 (CBS 614.71) designated acbA, B, C, D, E, F,
CC   G, H, I, J, K, L, M, N, O, P, Q, R, S, U, V, W, X, Y and Z, and asp3.1,
CC   3.2 and 3.3, and their homologues. The products of the invention
CC   individually or collectively, are used for synthesis or bioconversion of
CC   acarbose (or its precursors or related substances with alpha-glucosidase
CC   inhibiting activity), especially of alpha-glucosidase inhibitors. The
CC   products can also be used for optimising/inducing production of such
CC   compounds in Actinoplanes or other organisms. This sequence encodes a
CC   protein involved in the acarbose biosynthesis and bioconversion pathway
CC   which is described in the disclosure of the invention.
XX
SQ   Sequence 1350 BP; 176 A; 503 C; 482 G; 189 T; 0 other;

Query Match       10.8%;   Score 739.6;   DB 24;   Length 1350;
Best Local Similarity   73.4%;   Pred. No. 4.2e-83;
Matches   959;   Conservative   0;   Mismatches   344;   Indels   3;   Gaps   1;

Qy    739 GCCGTCTTCCAGCAGGCGGAGGCCCTCGTCGAGCAGCGCCGCCTCGTCGAGGGTGAACGGCGGT 798
          ||| |||||||||||||| ||||||   |||| | ||||||||| ||||||
Db   1319 GCCCGGTTCCAGCAGGTCCAGGCCGCGCAGCTCGGTCTCGTCGTCGATGACGAACGGCGGC 1260

Qy    799 GCCAGCCGCAGGATGTGCGCGCCAGGGAGTCGCAGCCCCAGGTCGCAGGGCGGTGGTG 858
          || |||| ||||||||||| |||||| ||||| ||||||||||| |||||||||||
Db   1259 GCGAGCCGGATGATGTGCCGCCGCCGACGCGGCGTTGCGCAGGACCGGCGGTGGTG 1200

Qy    859 TAGACGGCCCGGCGGTCGGGGGTGCCCGGCGGTCGGTGACGAACTCC 918
          ||||| ||||| | ||||| |||||||||||| |||||||||||||
Db   1199 TAGACCCGCGGGGATCTCCGGGGCCGGTCCGCGGTCGTCGCACGAACTCG 1140

Qy    919 AGGCCCCACAGCAGTCCGAGGCCCGGTACCTGCCGAGCTGGGAAGCCGGGACTCCAGG 978
          ||||| ||| |||||||||| ||||| |||||| |||
Db   1139 AGCCCGTGCAGCAGGCCGATGCCCGCCGACGTGCGCCCAGGTGCGGGTGCCGGTCGTGCATC 1080

Qy    979 GCGCGCAGCCGCCTCCTGGATGAGCTCGCCGGAGGACGCCCACGCGGTCGATCAGCCGGTCG 1038
          || | |||||||| ||  | |||||||| |||
Db   1079 TCGGCAGGCGGTCGGCGCAGCACCGCCGCGAGCTCCCGGACCTGCTCGATCAGCCCGGTCG 1020

Qy   1039 CGCTCGACGACCTCCAGCTGGCGCGGCGGCGCGGCGATCCCCAGTGGTTGCTCGCGTAC 1098
          |||| |||  |||||||||||||  |||||||||||| |||
Db   1019 CGGAGATCACCCAGAGCGGTGGCCGGCGGCCGGCGATCCCAGCGGGGCTGGCGGGGCGTAG 960

Qy   1099 GTCGAGGCGTACGCCCCGGGGTGGCGGCCGCCGCCTCCGGCCTGGCGCAGCTTCCCCG---CGTCCG 1155
          | ||||| | |||||||| |  | | ||||||| ||||
Db    959 GTGGAGGCGGTCGATCCGGCGAGCCCGGCCCGGGGATGCCGGGGATGCCGACCTCGTCACGCGCCCGG 900

Qy   1156 GCCAGCACGGGCGAAGGGAATCCGCTCCGGTGCCCTTGGACAGCATCGCCAGGTCCCGGC 1215
          | ||||||||||| ||| |||||| || | | ||||
Db    899 GCGAGCACGGCCAACGGGAAACCGGAAGCGGTGACCAGATCCGGG 840

Qy   1216 TCGATGCCGAACAGTTCGCTGCGGAGGAAGGCGCCGGTGCCGCCGGTGAGGACC 1275
          ||| |||||| |||||| ||||||||||||| ||||||||| ||||||||||||
Db    839 GGGACCCCGATCCCTCGGCTGCCAGAAAGGTTCCGGTCCGCAGGATCCGGTGAGCACC 780

Qy   1276 TCGGTCGGCGACGAGCAGCAGCACGCACCGCCGCCCCGGCGCGCCGGTCCCCAGCCTCCAGTAG 1335
          ||||||||| |||||| |||  | ||||| |||||| ||||||||| ||||||||||||||||
Db    779 TCGTCGGCGACCAGCAGTACGCGCTGTCCCGCCACACCGGCGGCCGGATCGCTCCCAGTAG 720

Qy   1336 CCCGGGGCGGCACGATCACGCGCCGGACGGTTCGAAGACCAGGGGCCCGAG 1395
          || ||||| ||||||| || | |||||||||||| ||||||||||| |||||
Db    719 CCCGGGTGGGGCGGCCAACGCGGCCGCCCAGCACCCGGTTCGAAGAAACAGGGGGCGTC 660

Qy   1396 ACGTTGGGCTTCTCCGCGATGTGCCGGCGACGAGGGTCGCCGCACCGCACGTCGCACGAG 1455
          ||||| ||| ||||||||||||||| |||| |||| |  | |||||||||||||||||||
Db    659 ACGTTGGGTTTCTCGGCCATGTGCCGGCGGTACCAGGGTGGCCATGTCGGCGACCTGCACGAC 600

RESULT 6
AAZ56004
ID   AAZ56004 standard; DNA; 3291 BP.
XX
AC   AAZ56004;
XX
DT   23-MAR-2000   (first entry)
XX
DE   Contig 003 from cosmid pKOS023-27 from Streptomyces venezuelae.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 003;
KW   ketolide; dNDP glucose synthase; dehydratase; picromycin; narbomycin;
KW   antibiotic production; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             104..982
FT                   /tag= a
FT                   /product= dNDP_glucose_synthase
FT                   /note= "glucose-1-phosphate thymidyl transferase"
FT   CDS             1114..2127
FT                   /tag= b
FT                   /product= dNDP_glucose4_6_dehydratase
FT   CDS             2124..3263
FT                   /tag= c
FT                   /product= PICCI
FT                   /transl_except= (Pos:2874..2876, aa:Xaa)
FT                   /note= "Xaa = Unknown"
XX
PN   WO9961599-A2.
XX

Qy   1456 GGGTACTTCCAGGCCCCAGGGGACAGCGGTAGCCAGTAGGGCTGTGTAGCCAGCACGCTGTTG 1515
          ||| ||||||||||||||| |||||| ||| || |||||||  | |||||||
Db    599 GGATAGCTCAGTCCAGGGCACCCGTAGCAGTAGTACGGGGCGTACCCAGCACGCGTGTTG 540

Qy   1516 CCGCTGAAGGCCTGGTGCGCGATGTCCCAGTCCCAGTGGGACCAGCATCCGGGCGCCCATGCGTCTTG 1575
          ||| ||||||||||||||||| |||||||||||||||  |||||||||||||| |||||
Db    539 CCGGCGAACGACTGGTCCCCCACGTCCACAGTGCACCAGCATCGGCGGCGCCCATCGTCTTG 480

Qy   1576 CCGTGGAAGCCGTGGCGCAGGGGCAGATCCGGTTCGCGCCGCGGTCGCCTGG 1635
          ||||||||||||||||||||||||||| || | |||||||||||||||||||
Db    479 CCGTGGAAGCCGTGCCGCAGTGCCCGACCCGGTTCGCGGCGGGTGCGGACGGCCTGG 420

Qy   1636 ACGACCCGCAGGGCGGCCGCTCGACCACCTCCGCCGCCGGTGAGAAGGCCTAGGTGTCG 1695
          ||||| || || || || |||||| || |||||||||||
Db    419 ACCGTCGCAGTGCCGCCCTCCACCACTCGCGCCGGCGCCCGGTGCTCGAGAGAAGGCGAGCGTGGTG 360

Qy   1696 AGCTGTTCGGCAGCCAGCAGTTCGGCAGCCAGTTCCAGCAGGCCGGTCCGGGTGGGCG 1755
          || ||||||||||||  ||||| |||||| |||
Db    359 AGGTGGTCGGGCAGCCAGTTCGGCAGCCAGAGTTCGCAGAGCGGCCGACCACCTCGGGCGGCGGTCGCGG 300

Qy   1756 CTGTCTGGACGTTCCACAGCGGCCGGCCCTGGGTGGTGAGTGCCTCGACGACCTCCGGG 1815
          ||||||||||||||||||||| |||||||| | |||| | ||
Db    299 AAGTCGTGGACGTTCCACAGCTTCCCGGCCTGTTCGGTGAGCCGGGCCACCACCTCGGGG 240

Qy   1816 TGCCCGTGCCCAGTGACTCGGTGAGGGTCCGGCGGCGGAAGTCGAAGTCAGGTACTGGTTGCCG 1875
          ||||||||||||||||||  |||||||||  ||||||||||||||||
Db    239 TGGCAGTGCCCGAGTGACTGGGTCAAAGTACCGGCGGCGGAGGTCGGAGGTATTCCACACCG 180

Qy   1876 TCCAGGTCGTCAGAAACGGAGACCGCGTCTCCTCGGGCAGGAAGACCCGGCCGTCCGTGGACGGCT 1935
          ||  || | | |||||| || ||||||| ||||  ||  |||||||| ||||||||
Db    179 TCGGCGTCGCCAGGGTCGCGGCCGGCCCGGCCCGCGACGAACACCCGCCGAGCCGCGCC 120

Qy   1936 TCCTCGGAGGCGCCCCGGCCAGGTGGCGGGCCTCCCGTGCCAGGTGCTGTCTCCGT 1995
          ||||| |||||  |||||| |||||||||||||||||| | | | |||| |||| |
Db    119 TCTTCGGAGGCTCCCGGAGGCCAGGTGACGGACTCCCGGGGGCGGCAGGTGGGCCTGCCTG 60

Qy   1996 AAGCCTGTCATCGCTGCCTCTGCTGCGGACCGGCGCGATC 2041
          |||| |||| ||| | | ||||||||| ||| |||
Db     59 CTCACCGGAACCGCCCGCCGGGTCGCCGCCGGTGTCGAAACGAAC 14

```
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67206, AAY67214, AAY67215.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Disclosure; Page 37-38; 98pp; English.
XX
CC   This is contig 003 from the recombinant cosmid pKOS023-27 DNA sequence
CC   (see AA256001) which contains a Streptomyces venezuelae DNA insert. The
CC   cosmid contains open reading frames which encode the various modules of
CC   the narbonolide polyketide synthase (PKS). The invention relates to
CC   recombinant DNA containing a coding sequence for a narbonolide PKS.
CC   Polyketides are compounds synthesised from 2-carbon units through a
CC   series of condensations and subsequent modifications. Modular PKSs are
CC   responsible for the production of many antibiotics including picromycin.
CC   The narbonolide PKS consists of a loading module, six extender modules,
CC   and two thioester domains. Four proteins make up the narbonolide PKS
CC   (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC   and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC   PICAIII includes extender module 5 and PICAIV includes extender module 6
CC   and a type II thioesterase domain. The second type II thioesterase
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid pKOS023-27. Narbonolide is desosaminylated in S.
CC   venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC   required for this conversion, and the desosamine biosynthetic genes are
CC   also found in cosmid pKOS023-27. The recombinant DNA of the invention is
CC   used to express, in transformed cells, narbonolide (or its derivatives)
CC   or other ketolides (particularly hybrids), which may then be converted
CC   (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC   polyketide antibiotics or their intermediates. The antibiotics are useful
CC   in human or veterinary medicine.
XX
SQ   Sequence 3291 BP; 521 A; 1289 C; 1044 G; 436 T; 1 other;

Query Match             7.8%;  Score 531.6;  Pred. No. 1.5e-57;
Best Local Similarity   53.4%;           DB 21;  Length 3291;
Matches 1477;    Conservative    0;  Mismatches 1179;  Indels 108;  Gaps 13;
```

```
Qy   2269 TGAAGGCCCTGGTCCTGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACACCG 2328
               |||||  |||||| || ||  || ||  || ||  || ||  || ||
Db    105 TGAAGGGAATAGTCCTGGCCCGGGGCGGGAGCGGAACTCCGGCTGCAACTCCGGGTCA 164

Qy   2329 CCGCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGGCTGCTCTTCTACGGCGTGAGTCC 2388
           | |  |||||| |  || || || | |||||| ||  |  ||  || |||||| ||  ||
Db    165 TTTCGAAGCAGATTCTTCCGGTCTACAACAAACCGATGATCTACTACTTCCGGCTGCGGTTC 224

Qy   2389 TCGGCGGGGGGTGTCCGGAGGGCCGGCGTGCGTCGTGGGCCGGTACGGCGCGGGAGATCC 2448
           || | |||| ||| || |  | | |||  |  ||  || | ||| || |  | |||| ||
Db    225 TCATGCTCGGCGGTATTCGGCAGATTCAAATCATCTCGGGTTACGGCATCACCTACCTCCACCAGC 284

Qy   2449 GCGA--ACTCACGGGACGACGG                                       2505
           | ||    || ||||| ||| |
Db    285 TCCAGTCGCTTCTCGGAAAACGGCAGGCACCGGCAGCGGAATAGAACTCGACTATGCGGTCCAGA 344

Qy   2506 CCCGCCCCGCTCCGTCCGCACGCACGGCGTTCGCCATCGGCCGCGGCTTCTGGGCGACGACG 2565
           |  | |  ||  ||||| | ||| ||||| |  || ||||| ||  || | || || || |
Db    345 AAGAGCCCGGAATCGCAGGAATCGGCGACGCGACGCACTTCTCGTCGGAGCCGAGCACGGCGACAGACA 404
```

```
Qy   2566 ACTTCCTGCTGTACCTGGGGGACAACTACCTGCCC---CAGGGCGGTCACCGACTTCGCCC 2622
              || |  |||| |||  ||||| |||||    |||   |||  || ||||| | ||||||
Db    405 CCTGCGCCCCTGATCCTGGGGCGACAACATCTTCCACGGCGGCCTTCTACACGGCTCCTGC 464

Qy   2623 GCCAATCGGCCGCGGCCGATCCCGCTGCGTCACCCGCTGCTCCTCGGCTACCGGTCAAGG 2682
             ||  |  |||| |  ||| | ||| ||  |  ||| |  ||  |||| || ||  |||
Db    465 GGGACAGCATCGCGGCGCCTGCGTGCGTCTTCGGCTGCGGGAACGTGCTGCCGTTGAAGGACCCCG 524

Qy   2683 CCGCCTTCGCGTCGCCGAGTTCGACGGCGGAACGTGCTGCCGTTGCGCTTCGAGAAGAAAC 2742
           | ||| | | |  ||||| || || || | |  ||  ||| ||  | || |||| || |
Db    525 AGCGGTACGGCGTCGCGAGTTGGACCGCCGACCGGCTGTACGGCTTCAGCCGCCGGCGTCC 584

Qy   2743 CCGACGTCCGCGCAGTCGCTCGCTCATCGGCGTGTACGGGTTCAGCCGCCGGCGTCC 2802
           || | || ||| | |  |  |  || | |  || | ||| | ||||| || ||| |
Db    585 CGTCAAGGCCGCTCCAACCTGCCGTCACCCGTCACCTCTACCTCTACGACAGCAGTCTG 644

Qy   2803 ACGAGGCCGTACGGCCATCACCCCCTCGCCGCGGCGTACGGGCCGA---GACCACCACCCGGCCCCT 2862
           ||| | || ||||||  | || || | || || | | || |||||   ||| ||| | ||||| ||
Db    645 TCGACATCGCCAAGAACATCCGGCCCTCGCCGCGGCGGCATCGGGAGTCGGGCTGGGGCTGCTGG 704

Qy   2863 TCGAGTCGGATCGAGGGGCTGCGCGGTACGGGCCGA---GACCACCACCCGGGCCCCT 2919
           | | || || | | ||| ||| ||||  |||| ||    || |||||| |||| |  |
Db    705 ACCGGCGTCTACCTGGAGCCGGCCGAACTCGTCAACCTGGGGCCCGGCGGCTTCGCCT 764

Qy   2920 GGGCGCGACACCGGCAGCCGCGGGAGGACATGCTGAGGTCAACCGTCAACCGCCACGGCGGAC 2979
           ||| | || | ||| | |||||  | |||| |||||| |||| ||||| ||||  |  |  ||
Db    765 GGCTGCACACCGGCCACCAGCTCCTCGGGCCCCAGCAGACGCGTGGTCGGCGGTCGGGGTCCGGG 824

Qy   2980 TGGAGGGCCCGCATCGAGGGGAAGGTCGACGCGCACGCGTCGGTCGGCGGTCGGGGTCGGCGCGGG 3039
                                                    ||| | || | | | ||| | |
Db    825 AGGAGCG---------------------------------GCAGGGCGTCTGGGATCGCGG 851

Qy   3040 TGGCCGAAGGCGGCGATCGTGCGGGGGTCACACGTGGTGGCCCGGTGGTGATCGGCGCGGG 3099
             || || |||||||||||| | ||| |||| || |||  |  || |  ||| | | | ||
Db    852 GCCTTGAGGAGATCGCCGTTCCATGGGCTTCATCGACGCCGAGGCCTGTCACGCGCCTGG 911

Qy   3100 GTGCCGTCGTCAGCAACTCCAGTGTGTCGGCCGTACACCTCCATCGGGGAGGAGGACTGCCGGG 3159
            ||||| | | ||| ||| || || |  || || ||| |  |   | ||| |||  | |  |||
Db    912 GAGAAGGCCTCTCCCGACCACCGAGTACGGCAGCTA------TCTGATGGAATCGCCGGGGGGT 964

Qy   3160 TCGAGGACAGCGCCATCGAGTACTCCGTCGTGCTCCGTGCGGCGCCCCAGGTCGGAGGGGGCGGT 3219
           |  |||| || |||| ||||  |   ||  | |  || | || | || || | ||| || |  |
Db    965 GCCGGAGGGAGCCCGTGAGGGCACTCGGCGGCCACTCGGCACCGGACCGACAGCCGCCA 1024

Qy   3220 CCCGCATCGAGGCGTCCCTCATCGCGCGACCCGCCGGCTCGTCGGCCCCCGTCTCC---------- 3279
           ||| || | ||| || | | | | | || || ||||| | |  |  || || |||
Db   1025 CCGACAGTGCGACCACCACCACGCCGACCAAGGTGTATCTCACCCCATGACCAC 1074

Qy   3280 CGCAGGTCACCGACTGGTGATCGGACCTACAGCGCGACCAAGGTGTATCTCACCCCATGACCAC 3339
                            ||||| | |||  | || |||| | |||  |
Db   1075 ------CACACCGCGACCTACGCCGAGCAGGCGACCGAAAGGAAG----------ACGGCAGT 1115

Qy   3340 GACCATCCTGTCACCGGCGAGTTCCGCGGCTTCATTCGGTGGCGGACCTGGCCGCACTCGGTA 3399
           | || |  ||| ||| | || ||   | |||  | || ||| || | || ||| |||| | |
Db   1116 GCGGCTTCTGGTGGTGACCGGAGGTGGCGGCTTCATCGGTGACCGTGACCTCGCAGCCTCCTA 1175

Qy   3400 GTCGCCGGGGCGCCCCCG-----GGGCGGTCTCCCGGACGTGCCGCGCTGACCGGCGACCTCGGTCA 3453
           |  |  | |||  | |||     ||| || |  || ||| |  || ||  | |  ||  || | ||
Db   1176 CGCCGGCAGCCTCGACCTCGGGTCTACCCGCTGCTCGACGTGCTGACCATCCGCTGCCTCACCTA 1235

Qy   3454 CGCCGGCAGCCTCGACCTCGGGGCCGTCGCGCTGACCATCCGCTGCCTCACCTTCGTCCA 3513
           ||||| | || || | ||| || |  | ||| || | ||||| || |  | |  |  ||
Db   1236 CGCGGCAACCGGCAACCTGCGGCTCGTCGGGCGTCCGGGACGCCGACTCGGCGTTCGTCCA 1295

Qy   3514 GGGCGACGTGTCGCAGATTCAAATCATCTCGGGTTACGGCATCACCTACCTCCACCAGC 3573
           ||| | || ||  | | | |  | || | ||| ||| | |   | ||   |  ||| |
Db   1296 CGGCGACATCCGGCGCACGCGGCTCCCCGGAACTGCGGGCCGTTCATCGACCTGAACTCGT 1355

Qy   3574 GCACTTCGGCGCCAGTCGACTCGACTCGACCGGCTTCTCGGGCGACGACG 3633
           ||||| || | || || ||| ||| ||    || ||| ||| || ||||
Db   1356 CCACTTCGGGCGCCAGCAGCCACGGTGACCGGTGACCCTCCGTGTTCACCGA 1415

Qy   3634 CACCCAACGTCTGGGCACGGGCCGCGGCAGCAGGAATCGCAGGAATCGGCGACGCGACGACA 3693
```

RESULT 7
AAA75636
ID    AAA75636 standard; DNA; 3292 BP.
XX
AC
XX
DT    22-JAN-2001  (first entry)
XX
DE    Nucleotide sequence of ORF14 which encodes dNDP-glucose 4,6-dehydratase.
XX
KW    Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW    antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KW    desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW    picromycin biosynthesis; ss.
XX
OS    Streptomyces venezuelae.
XX
PN    US6117659-A.
XX
PD    12-SEP-2000.
XX
PF    27-MAY-1999;        99US-0320878.
XX
PR    28-MAY-1998;        98US-0087080.
PR    22-SEP-1998;        98US-0100880.
PR    08-FEB-1999;        99US-0119139.
PR    20-MAY-1999;        99US-0134990.
PR    30-APR-1997;        97US-0846247.
PR    06-MAY-1998;        98US-0073538.
PR    28-AUG-1998;        98US-0141908.
XX
PA    (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI    Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR    WPI; 2000-610844/58.
XX
PT    New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT    for converting ketolides to antibiotics and as antibiotics and
PT    intermediates in the synthesis of compounds with pharmaceutical value
-
XX
PS    Disclosure; Columns 43-46; 117pp; English.
XX
CC    The present sequence is used to produce the recombinant DNA compounds
CC    of the invention. The specification describes a recombinant DNA compound
CC    expressing recombinant polyketide synthase genes in host cells for the
CC    production of narbonolide, narbonolide derivatives and polyketides that
CC    are useful as antibiotics and as intermediates in the synthesis of
CC    compounds with pharmaceutical value. The DNA compounds may also encode

```
CC    a C12-hydroxylase (pickK), desosamine biosynthesis and desosaminyl
CC    transferase enzymes (useful for conversion of ketolides to antibiotics),
CC    and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC    These compounds are also useful for increasing the antibiotic activity
CC    of a compound relative to the unhydroxylated compound. The recombinant
CC    host cells are useful as genetic systems that allow rapid engineering
CC    of the narbonolide polyketide synthase. These would be valuable for
CC    creating novel ketolide analogs for pharmaceutical applications.
XX
SQ    Sequence 3292 BP; 521 A; 1290 C; 1044 G; 436 T; 1 other;

Query Match            7.7%;   Score 529;  DB 21;  Length 3292;
Best Local Similarity 53.4%;   Pred. No. 3.1e-57;
Matches 1476;  Conservative     0;  Mismatches 1180;  Indels 109;  Gaps  13;
```

```
Qy   2269 TGAAGGCCCTGTCCTGGCAGTGAACCGGCAGCAGACTGAGGCCGTTCACCCACACCG 2328
          ||||||  | | ||| ||  | ||| ||||||||| |   | |   ||
Db    105 TGAAGGGAATAGTCCTCGGCCGCGGGAGCGGAACTCGGCTGCCATCCGGCGACCTCGGTCA 164

Qy   2329 CCGCCAAGCAGCTGCTCCCCATCGCCACAAGCCCGTGCTCTTCTACGGCCTGGAGTCCC 2388
          |  |||||  | |   |||  || |||||  | ||   |||| |||||
Db    165 TTTCGAAGCAGATTCTTCCGGTCTACAACAAACCGATGATCTCACTACTATCCGGTCGGTTC 224

Qy   2389 TCGCGCGCGGGGTGTCCGGGAGGCCCGGCGTCGTCGTTGGGCGCGTACGGCGGGAGATCC 2448
          ||   | || ||  || | |  | | | | || | |||||  |
Db    225 TCATGCTCGGCGGTATTCGCGAGATTCAAATCATCTCGACCCCCAGCACATCGAACTCT 284

Qy   2449 GCGA---ACTCACCGCGACGGCACCGCGTTCGGGGTTACGGCACCTACCTCCACCAGG 2505
          ||  |    ||   ||  || || |  |||||  || |||||
Db    285 TCCAGTCGCTTCTCGGAAACGGCCAGGCACCTGGGAATAGAACTCGACTATGCGGTCCAGA 344

Qy   2506 CCCGCCCGCTCGGTCGCGGTCACCGCATCGGCCGTTCCTGGCCGACGACG 2565
          ||   | | |||    ||| ||| |  | ||  |
Db    345 AAGAGCCCCAGGAATCGCGGCCGCACTTCTCTCGCCACAATCGGCGACGCACTCGCGACGCA 404

Qy   2566 ACTTCCTGCTGTACTCGCCCTGCCC---CAGGCGCGTCACCGACTTCGCCC 2622
          ||  |  | |||| |||  |||    |||  | | ||  ||||| ||
Db    405 CCTGCGCCCCTGATCCTGGCGCGACAACATCTTCGCCGGCCCCGGTCTCACCGGTCGCGGTTC 464

Qy   2623 GCCAATCGGCCGCCGATCCCGCGGCGGGCCCGGCCTGCTGCCCGGTGTCACCCGGACCCGT 2682
          ||   || |  |  |||||  |||||  |
Db    465 GGGACAGCATCGCGGCGCCTCGACGGTGCCGCGAAGGGAAGCTGCTGGCTCCGTCGCGCCCG 524

Qy   2683 CCGCCTTCGGCGTCGGAGGTCGCGACGGGAACCTGCTGGACCGGCCTCGTCGAGAGAGC 2742
          |  |||||  ||  |||| |||| |  ||  ||||||  |
Db    525 AGCGGTACGGCGCTCGCCAGGTGGACGCGACGGGCGACGCGGGGGCTCGCCAGGCTGAGAGC 584

Qy   2743 CCGAGTCCCGCCGCAGCTCGCTCGGCCTGCCTCGCGCCCATCACCCCTCCGCCCGGTCC 2802
          ||   |  | ||  || |||  ||  |||||  ||
Db    585 CCGTCAAGCCCGCGTCGCGCTCGCCGGCCTCCAACCTCGTACCCGCGGGAGCTCGTGAGAGC 644

Qy   2803 ACGAGGCCGTACGGGCATCACCCCGCATCACCCGCGGGCACCGCCGGAGCTGGAGATCACCACGCCG 2862
          |||  ||  || |||  | | |  || ||||||||
Db    645 TCGACATCGCCAAGAACATCCGGGCCCGGCTCCAACCTCGTGAGGTCAACTGGTCCGAGATCGTCA 704

Qy   2863 CCGACAGTGATGATCGACCGGGCCTGCGCGTACGGGGCCGGA---GACCACCCGGCCCGCCCCT 2919
          ||   ||  ||  |||  |  |||||| ||| |||  |||||    |||  ||  |  |||
Db    705 TGCAGTGATGATCGACCGGGCGCTGCCCGGGGGGCTGGGGATGGCCGTCACGTCGGGGTCCGGG 764

Qy   2920 GGCGGGACCGCAGCGGCCGGAGGACTCGACGGGCGCACAGCACCGGGGCTGGTCGGTCCGGG 2979
          |  |||  ||  ||||| ||  ||  ||| ||
Db    765 GGCTGGACACCGGCACCGCCACCCACGCGGGGTCACACGTGGCGTGGGGCGCAGGGGTGGAAC 824

Qy   2980 TGGAGGGCGCCATCGAGGGGAAGGTCGACGCGCACAGCGTCGGGGCCCGGCGGTCCGGG 3039
          |  || |||                       |  ||
Db    825 AGGAGCG------------GCAGGGCGGTGCCGGTGTGATCGGCGCGG 851

Qy   3040 TGGCCCAAGGCGCCGATCGTGCTGGCGGGTCACACCTGGTGGCCGGTGATCGGCGCGG 3099
          ||    ||   ||  ||  | |||
Db    852 GCCTTGAGGAGATCGCCTTCGCATGGCTTCATCGACGCCGATGGCGCGGGAGGACTTCGCC 911

Qy   3100 GTGCCCTCGTCAGCAACTCCAGTGTCGCCCGTACACCTCACGTCGCCCATCGGGGAC 3159
          | | |||  ||
Db   3100 ...
```

```
Db    912 GAGAAGGCCTCTCCGCCACCGAGTACGGCCAGCTA------TCTGATCGGAGATCGCCGGC 964

Qy   3160 TCGAGGACAGCGCCATCGAGTACTCCGTCCTGTCCGTGGCGGCCGGGGGCGGGGGGCGGT 3219

Db    965 CGCGAGGGGAGCCCCGTGAGGGAACCGCACCTCGGCGCGGCCGCCGGCTTCCCACGACCGCAGCGCCA 1024

Qy   3220 CCCGCATCGAGGCGTCCTCATCGGCGCCGGCACCGGCGACCGCACCGCACCGCACGACCGCAGTGCGACC------ 1074

Db   1025 CCGACAGTGCGACCACCACCGCGACCGCACCACCACCCAAGGTGTATCTCACCCATGACCAC 1074

Qy   3280 CGCAGGCTCACCGACTGGTGATCGGCGACCAGCAAGTGTATCTCACCCATGACCAC 3339

Db   1075 -----CACACCGCGACCTACAGCGCGACGCGAAAGGAAG----------ACGGCAGT 1115

Qy   3340 GACCATCCTGTCACCGGAGCCGGAGCGGCGCTTCATTCGTCCGGGCTTCCGTCGCGGCTCCT 3399

Db   1116 GGGGCTTCTGGTGACCGGAGGTGCGGGCTTCATGCGGCTTCCTGCGGGCAGCTCCT 1175

Qy   3400 GTCCCGGGGGGGCCCG------GCGGCGTGCGGGTGACCGGTGACCGGTGCGGTCA 3453

Db   1176 CGCCGGGGCGTACCCCGACGTGCCGCCTGCACGGGTGCGTGACCATCCGGCCTCACCTCGTCA 1235

Qy   3454 CGCCGCAGCCTGGGCCGTCGCTGCGCTGACACGCGTCGGCCGGCGGGGCGGGACCGCGGAGATCC 1295

Db   1236 CGGGGCAACCGCGCCAACCTCGCCCGTGGACGCGGCGACCCCGGACTCGGCGCTTCGTCCA 1295

Qy   3514 GGGGACGTGTGGACACCGCGGTCGTCGACACGCGCTGGCCGCGGCGGGCTGCGGGCATCGT 3573

Db   1296 CGGGGACATCCGGACGCGCGGCCTCGTCCTCGCGGAACTGGCGGTGCCCTTCACCCG 1355

Qy   3574 GCACTTCGGGCGCCGAGTCGCACGTCGACGCGCTCGCACGGCGGCCGTCCGTGTTCACCGA 3633

Db   1356 CCACTTCGCGGGCGTCGCGCGAGAGCCACGTGGACCGCTGCTGCTGGCTCCTGCACGCGGTGCCCGAC 1415

Qy   3634 CACCAAGCTGTGGGCACCCAGGTCCTGCT-GCACGGCGTCCAGTGCCGGTCGGCGGTCGGCCGG 3693

Db   1416 GACCAAGCTGCAGGGCACCGCAGACGCTGTACGGGTCCGCCCCGCGGA 1475

Qy   3694 CTTCGTGCAGTCTCCACCGACGGAGGTGTACGGGTCTACGGGTGACCGGTCGCGGTCTGGACCGA 1535

Db   1476 GGTCGTGCAGTCTCCACCGACGAGCGTGTACGGCGTACGGGTCGCGGTCTCGAAGGCGG 1535

Qy   3754 GAGCGCGACCCCTGCTTCCGACCTCGGCCGTGCTGTCGAAGGCCGTCGCGGGTCTCGGACCCTCAT 3813

Db   1536 GAGCAGCCGTGGACCGTGGCCTGCTGGAGCCCAACTCGCCCTACGCGCCCTACGGCTCCGACCTCGT 1595

Qy   3814 GGCGGTCGCGCCACCAGCTCCAGCAGCGCCAACCTCGTCGAGAGC 3873

Db   1596 TGCCCGGGCCGTACCAGCCCTACCGGCCTCGACCTACGGGTCCTGGACCGA 1655

Qy   3874 CTTCGGCCCCCAGCGCATCCCGAGAGCTCATACGGCGGGCCTTCCTGACCGAGCCTCCTGTC 3933

Db   1656 CTACGGGCGTACCAGCCACCCAGCACCAGCACCTTCTTCGTGACGGACAACCTCTCGA 1715

Qy   3934 CGGCGGCACCGTTCCCGTCCGGGGCCATCGCGCGGCATCGCGGGCCGGCTCGCGGACTGCACGTCGA 3993

Db   1716 CGGCGGGGACGTCCCGGTTCGTACGGGCCGGGCCTGCGGGCCGGGCCGGGCGAGATCTCCGGGTGGGTCACACCA 1775

Qy   3994 CGACCACGTTCAGCGGCCGGAGCCTGTCGAACTCGTCCGCGGCTCGAGGAGAGGATCTACAA 4053

Db   1776 CGACCACTGCGGGCCGCATCGGGCTGTGGACTCGTCGACTTCCAGTCGGTCCGGG 1835

Qy   4054 CATCGGGGGCCGGCCCGGAGCGCATCTCGCTGCGAACTCGTCCGAGCTCCTGACTCTACCA 4113

Db   1836 CATCGGGGGACTGTCCGGCCGTGGTCCTGGAGCTGTCCAGCTCTCCTGACTGGCT 1895

Qy   4114 CGGCGGCCCGGAGTCGGCCATCTCCGGTCGGTGTGTATCGGCGGCGG 4173

Db   1896 CGGGGGCCGGCCGACTGGTCTCCGGTCGCGGTCGACCGAAGGGTCGCCACCGACCTCGGC 1955

Qy   4174 CGGGTCGACCACCACAGCGCAGATCGAGGCGGGACTGCGGGGGACTCGGGGAGGACTGCCGGG 4233

Db   1956 CTCCCTTCGTCGAGCAACTCCAGGTGCGGGACACTGCGTACACCTCCATCGGGGTCTCCCTTCGC 2015
```

XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAV77179.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin .
XX
PS   Claim 2; Page 281-287; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 other;

Query Match             7.7%;  Score 527.4;  DB 21;  Length 12441;
Best Local Similarity   53.3%;        Pred. No. 4e-57;
Matches 1475;  Conservative  0;  Mismatches 1181;  Indels 109;  Gaps  13;

[sequence alignment data]

RESULT 8
AA287284/c
ID   AA287284 standard; DNA; 12441 BP.
XX
AC   AA287284;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae desosamine biosynthetic gene cluster plkB, SEQ ID NO:3.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.

```
Qy   4693  GTGCAGCCCAGGTGCGATCAGCGGGACGCGG-CGGCCTGGCCCCCTCCAGGAGAACAG  4751
                 |-|-|||-|||-| |-||| |||-||| ||| |||-|| |||||| |||| ||
Db   6910  GAGAAGGCGATCACCCCCGCACCCGGGCGCGTCCTCCCCGTCCACCTTACGGGCCACCCC  6851

Qy   4752  CGAGTGCTGCACGAGCTCCTCGGACTCCCGCCGCGACACTCCCAGGTGCTCCCGCACGCC  4811
                 |-|| |||| |||| | |||-||-||| ||| ||-|| |-| |-|||
Db   6850  GCCGACATGGACGCCCCTCCGGAGCTCGCGGACGGCCTGCACATCGTCGAGGAC  6791

Qy   4812  GGCCCGGAACCCTCGATCTTCCGGCTGCACCGGCCACGAAGCGGGCGGGCCCGACGGCGAG  4871
                 ||||| ||||| |||-| |||-|| || ||||  ||| |-||  |||-||| |||
Db   6790  GCCGCGCAGGCCCACG--GCGCCCGCTACCCGGGCCGGCGGGATCGGCGCCGGGGGTCGTCG  6734

Qy   4872  GCCGACGGCGGCTCGCTGCCCCAGTCCGCCAGGTGCGCCACGGCCAGGCGCATCGCGGCCCG  4931
                 |-||||| |||-||  |-||   |||-|| |||-||-|  |-| |||||||-|
Db   6733  GTGGCCCGCGTTCAGCTTCTACCCGGCAAGAAGCCTCGCTGCTTCGGCGACGGCGGCGCC  6674

Qy   4932  GTCTCCGGGGAGACGAAAGGGTGCCTCGATCCGGGGCGAGAAACCCGTTCACGAGGACGAA  4991
                 |||||| ||  |-| ||| | ||||-|||  |||  |||||| ||
Db   6673  GTCGTCACCGGCGACCCCCGAGCTCGCCGAACGGCTCCGGATGCTCCGGAACTACGGCTCG  6614

Qy   4992  GGGCA  4996
                 ||||
Db   6613  CGGCA  6609
```

RESULT 9
AAZ87319/c
ID   AAZ87319 standard; DNA; 13613 BP.
XX
AC   AAZ87319;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae desosamine biosynthetic gene cluster pikB.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..809
FT                  /*tag= a
FT                  /partial
FT                  /product= "PikB gene cluster protein #1 (AAY77204)"
FT                  /note="No initiation codon given in the specification"
FT   CDS            806..2014
FT                  /*tag= b
FT                  /product= "PikB gene cluster protein #2 (AAY80998)"
FT   CDS            2162..6741
FT                  /*tag= c
FT                  /product= "PikB gene cluster protein #3 (AAY77205)"
FT   CDS            6834..7402
FT                  /*tag= d
FT                  /product= "PikB gene cluster protein #4 (AAY77206)"
FT                  /transl_except= (pos:6837..6841, aa:Gln)
FT   CDS            7492..8205
FT                  /*tag= e
FT                  /product= "PikB gene cluster protein #5 (AAY77207)"
FT                  complement (7942..8205)
FT   CDS            /partial
FT                  /*tag= f
FT                  /product= "PikB gene cluster protein #6 (AAY77208)"
FT                  /note="No termination codon given in the specification"
FT                  /transl_except= (pos:8270..8272, aa:Val)
FT                  /transl_except= (pos:8273..8275, aa:Thr)
FT                  /transl_except= (pos:8276..8278, aa:Gly)
FT   CDS            complement (10126..11139)
FT                  /*tag= g
```

FT   CDS            /product= "PikB gene cluster protein #7 (AAY80999)"
FT                  complement (11271..12149)
FT                  /*tag= h
FT   CDS            /product= "PikB gene cluster protein #8 (AAY77209)"
FT                  complement (12342..13799)
FT                  /*tag= i
FT   CDS            /product= "PikB gene cluster protein #9 (AAY77210)"
FT                  complement (13706..15043)
FT                  /*tag= j
FT   CDS            /product= "PikB gene cluster protein #10 (AAY77211)"
FT                  15404..15574
FT                  /*tag= k
FT                  /product= "PikB gene cluster protein #11 (AAY77212)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;     99WO-US14398.
XX
PR   26-JUN-1998;     98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77204, AAY77205, AAY77206, AAY77207, AAY77208,
DR   AAY77209, AAY77210, AAY77211, AAY77212, AAY80998, AAY80999.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 32; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 32.
XX
SQ   Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;

Query Match             7.7%;  Score 527.4;  DB 21;  Length 13613;
Best Local Similarity  53.3%;  Pred. No. 4e-57;
Matches 1475;  Conservative    0;  Mismatches 1181;  Indels 109;  Gaps  13;

Qy   2269  TGAAGGCCCTGGTCCTGGCAGGTGGAACCGGCAGCAGCAGACGCGTTCACCCACCACCG  2328
                 ||||||  |  ||-|||||| || ||-||| |-||| |||-|| |||  ||| |||
Db   9988  TGAAGGGAATAGTCCCTGGCCGCCGGGCGGGCGAAGTCGGCTGCATCCGGCGGACCTCGGTCA  9929

us-09-922-683-7.rng

Thu May 29 15:11:29 2003

Qy  2329 CGGCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTCACGCGTGGAGTCCC 2388
         ||||   ||| || ||||   || ||| || |||| ||  | ||| | || ||
Db  9928 TTTCGGAAGCAGATTCTTCCGGTGTTCCGGAGGCCGGCGTCGTGGGCGTACGGCCGGGAGATCC 9869

Qy  2389 TCGCCGCGGCGGGTGTCGGGTGTCGTCGTGGGCGCGGGGAGATTCAAATCATCTCGCAGACCATCCAGCACAATCGAACTCT 9809

Db  9868 TCATGCTCGGCCGGTATTCGCGGAGATTCAAATCATCTCGGTTAGGCATCACCTACCTCCACCAGC 8738

Qy  2449 GCGA---ACTCACCGGCGACGGCAACGGCACCGGTCGGCATCGCCCCGGCGTTCCTGGCGACGACG 2505

Db  9808 TCCAGTGCTTCTCGGAAACCGGCAGCCAGCCACTGGAATAGAACTCGACTATGCGTCCAGA 9749

Qy  2506 CCCGCCCGCTCGGTCTCGCGCACGGCGGTGCGCATCGCCCCGGCGTTCCTGGCGACGACG 2565

Db  9748 AAGAGCCCGCAGGAAATCGGCGGCGCACTTCGTCGGCAGCCGAGCACATCGGCGACGACA 9689

Qy  2566 ACTTCCTGCTGTACCTGGGGACAACTACCTGCCC---CAGGGCGGTCACCGACTTCGCCC 2622

Db  9688 CCTGCGCCCGATCCTGGGCGACAACATCTTCCACGGCCCTCTACACGCGCTCCTGC 9629

Qy  2623 GCCAATCGCGCGCGCGATCCCGCGGCTGCTGCCACCCCGGTCGCCGGACCCGT 2682

Db  9628 GGGACAGCAGCATCGCGGCGCTCGCGGCGCTCGTGCTCTTCGGGTTACCCGGTACCCGGTCGCTCCGGGCTCAAGGACCCCG 9569

Qy  2683 CCCGGCCTTCGGCGCGTCGCGGAGGTCGACGGCGAACGTGCTGCGCTTGACCGACGAAAC 2742

Db  9568 AGCGGGTACGGCGTCGCCGAGGTGGACGGCGACGGGCCCCGGCTGACCGTGTCGAGAAGC 9509

Qy  2743 CCGACGTCCCGCGAGCTCGCTCGCGCCGAGCTCCTTCAGCGCCGGTCGGCCGTCC 2802

Db  9508 CCGTCAAGCCGGCCGCGCTCCAACCTCGCCGTCACCGGTCGTCACCGGCATCACGACGCTGC 9449

Qy  2803 ACGAGGCGGTACGGGCGCGATCCACCCCCGCGCCCCTCGCGCCCTCGCCGCTCCACCGAGCTCGGAGATCACCACCG 2862

Db  9448 TCGACATGCGCCAAGACACATCCGGCGCCCTGCCCCTACCGGCCCGCGAGAGTCGGAGATCACCACCGACGACTCA 9389

Qy  2863 TGCAGTGGATGATCGACGCGGCGCCGTTGCGCGGGCCGCCGA---GACCACCACCCGGCCCT 2919

Db  9388 ACCGGCGTCTACCTGGAGCGGGGCGGGCCGGGCCGGCGTTCAACCTGGGCCGGCTTCGACGGAC 9329

Qy  2920 GGCGGGCACCGCAGCGCGCGAGAGACACTCGGCGCCAGCGCCACCACCACTCGGCGCGGCCGCAGCGCCA 2979

Db  9328 GGCTGACCACCGGCGACCGGCCGCCTGGGGCCATCGGGGGGCCCGCCGT 9269

Qy  2980 TGGAGGGGCCGCCATCGAGGGGAAGGTCGACGCGGCACGCTGTGACGGGGGTCGGCCGTCGGGG 3039

Db  9268 AGGAGCG--------------------------------------GCAGGGGCGTCGTGGATCGGCGG 9242

Qy  3040 TGGCCGAAGGCGCGATCCGCGCGGGGTCACACGTGTGGGCGGTGGCGTGCGCGGCCTGG 3099

Db  9241 GCCTTGAGGGAGATCGCCTTCGCGGCGGGGGCGGGGCCGCATCGGGGCAGCCGGAGGCTGAGTCGGACGGAC 9182

Qy  3100 GTGTGGGTCGTCAGCAACTCCAGTGTCGCGGGCACGAGTACGGCAGCTA------TCTGATGGAGATCGCCGGC 9129

Db  9181 GAGAAGGCCTCTCCGACCCTCGGACCACTCGAGTACGGCAGCTA------TCTGATGGAGATCGCCGGC 9069

Qy  3160 TCGAGGACAGCCGCCATCGAGTACTCCGGCCGGCACCTCGCCGCCCGCGCGCGCCCCGTCC 3219

Db  9128 CGGCGGGAGCCCGCCGCCGTCCCTCGAGGGGCCACCGCCCGCGGCCGCCGCGTCTCGCCGCGCTGTCGGCATCGCCGACCA 9069

Qy  3220 CCCGGCCATCGAGGCGTCCGTCGCGCCCGCCGACCGTGCGCCGCCCCCGTCC 3279

Db  9068 CCGACAGTGCGACCCCGCCCCCGGCGGCGGCCACACCCGGCGACCAGTGCGACCAC 9019

Qy  3280 CGGCAGGCTCACCGACTGGTGATCGGCGACCAAGGTGTATCTCACCCCATGACCAC 3339

Db  9018 -----CACACCGCCACCGCGACCGGGGGCGGGGGG------------ACGGCAGT 8978

Qy  3340 GACCAATCCTCGTCACCGGCGCTCAATTCGCTCGCCTGCACTTCGTCGCGGCAGCTCCT 3399

Db  8977 GCGGCTTCTGCGTGGTCGGTGGCCGGAGGTGCGGGGTTCCGGCGGCGCCGGGG------ 8918

Qy  3400 GTCGCGCCGGGGCCCCG------GCGGCGTCGCGCGGCGTCGTGATCGCCGGCGTCGGCGAGTCGGACGGCGT 3453

```
Db  7837 CTCGAAGGATTCGAGGCGGAGTTCGCCGCGTACTGCGAGACGGACCACGCCGTCGGCGTG 7778
              ||||||  ||| ||         |  ||| ||||| |  |  |||| || || |||||||
Qy  4521 CCCCATCGCCTTGACGGGCT------GTCTGATGGTGGTCAGGGGAGGTCGGTGAAG 4572
Db  7777 AACAGCGGGATGACGGCCCTCCAGTCCCCTCCGGCCTCCGGCATCGGACCCGGGGAC 7718
             |  |||  || || ||     ||||||| | |||| ||| ||  ||  |||| |||||
Qy  4573 GCCATGAGCGGCGAGTCGTCGAAGCCGACCACCGAGATGTCACCGGTGAGACCC 4632
Db  7717 GAGGTGATCGTCCCCTCCACACGTACATCGCCAGCTCAGCCTCCGGGTCCGCCACCGGC 7658
             |  || || ||     |  | || ||| |||| | || |||  |||| |||||| |  |
Qy  4633 CGCCGGCGCGCCGCACGCGCCGAGGCGCCAGGGCCATCATGTCGCTGGCCCACATGACGGCG 4692
Db  7657 GGGACCCCCGTGCCCGTCGAGCCGCACGAGGACCACCCACCCTGGACCCGTGCTCGTC 7598
             | || |||||  || | ||||| ||   ||      |||| |||| ||  ||||
Qy  4693 GTGCAGCCCAGTCGATCAGCGCGGGACGCGG-CGGCCTGGCCCCCCTCCAGGGGAGAACAG 4751
Db  7597 GAGAAGGGGCATCACCCCCGACCTCCTCCCCGTCCTCCACCTCTACGGGCACCCCC 7538
             ||  || |  |  |  |||| | || || ||||| | | |||  ||||| ||
Qy  4752 CGAGTGTGCAGAGCTCCTCCGGACTCCCGCGCACACTCCCCAGGTGCTCCCCGCACGCC 4811
Db  7537 GCCGACATGGACGGCCCCTCCGGAGCTCGCGGACGCCGCACGGCCGCACATCGTCGAGGAC 7478
             | ||     ||     |||  |    ||| |||  | | || |||    |||  || |||
Qy  4812 GGGCCGGAACCCCTCCGATCTTCCGCTCCACCGGACGCCGCGAGACGGGCCCGACGGGCGAG 4871
Db  4477 GCGCCGCAGGCCCACG---GGCGCCGGTACCCGGGGCCCGGCGGATCCGGGGTCGTCG 7421
             |  ||||||  |          |||  |   |||||  || ||  |    ||||| |||
Qy  4872 GCCGACGCGGCTCGTGCCCCAGCTGCGGCCAGGTGCGCCCAGGGCCATCCGGGGCCCG 4931
Db  7420 GTGGCCGCGGTTCAGCTTCTACCCCGGCCAAGAACCTCGGCTGCTTCGGCGGCGGCCGCC 7361
             |  |||| ||   |  | ||  |    ||    |||  || ||   |  || || ||||
Qy  4932 GTCGTCCGGGAGACGAAGAGGGTGCCTCGATCCGGGGCGAGAACCCGTTCACGAGGACGAA 4991
Db  7360 GTCGTCACCGGCGCACCCCGAGCTCCGAACGGCGTCCGCAACGGCTCCGGAACTACGGCTCG 7301
             ||| | ||||| |  | ||||  ||  | | || ||  |  ||| ||  ||||||||  ||
Qy  4992 GGGCA 4996
             |||||
Db  7300 CGGCA 7296
```

RESULT 10
AAD39043/c

```
ID   AAD39043 standard; DNA; 13613 BP.
XX
AC   AAD39043;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Streptomyces venezuelae desosamine gene cluster.
XX
KW   Glycosylated polyketide; modified recombinant bacterial host cell;
KW   mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
KW   tetracycline; polyene; polyether; ansamycin; isochromanequinone; sugar;
KW   desosamine; des; gene; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..809
FT                   /*tag= a
FT                   /product= "Streptomyces venezuelae Des gene cluster
FT                   encoded protein #1"
FT   CDS             806..2014
FT                   /*tag= b
FT                   /product= "Streptomyces venezuelae Des VIII protein"
FT                   /transl_except= (pos:806..808, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             2162..3442
FT                   /*tag= c
FT                   /product= "Streptomyces venezuelae Des VII protein"
FT   CDS             3535..4245
FT                   /*tag= d
FT                   /product= "Streptomyces venezuelae Des VI protein"
FT                   /transl_except= (pos:3532..3534, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             4312..6741
FT                   /*tag= e
FT                   /product= "Streptomyces venezuelae Des R protein"
FT                   /transl_except= (pos:4312..4314, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (7969..6829)
FT                   /*tag= f
FT                   /product= "Streptomyces venezuelae Des V protein"
FT                   /transl_except= (pos:7969..7967, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (8979..7966)
FT                   /*tag= g
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /transl_except= (pos:8979..8977, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (9111..9989)
FT                   /*tag= h
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT   CDS             complement (11639..10182)
FT                   /*tag= i
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT   CDS             complement (12883..11636)
FT                   /*tag= j
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /transl_except= (pos:12881..12883, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO20029035-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31255.
XX
PR   05-OCT-2000; 2000US-238185P.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (LIUH/) LIU H.
PA   (SHER/) SHERMAN D H.
PA   (ZHAO/) ZHAO L.
XX
PI   Liu H,  Sherman DH,  Zhao L;
XX
DR   WPI; 2002-405171/43.
DR   P-PSDB; AAE24228, AAE24229, AAE24230, AAE24231, AAE24232, AAE24233,
DR   AAE24234, AAE24235, AAE24236, AAE24237, AAE24347.
XX
PT   Modified recombinant bacterial host cells in which the expression and
PT   activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT   altered, useful for producing metabolites with altered sugar structures
XX
PS   Disclosure; Fig 8; 174pp; English.
XX
CC   The invention provides a method to alter the sugar structure diversity
CC   for a particular metabolite via the recruitment and collaborative action
CC   of sugar genes from a variety of sugar biosynthetic pathways to yield a
CC   metabolite comprising a non-natural sugar, e.g., a novel glycosylated
CC   polyketide. The invention also relates to a modified recombinant
CC   bacterial host cell (mRBHC) in which the expression and activity of
CC   nucleic acids encoding sugar biosynthetic enzymes has been altered.
CC   The mRBHCs may be cultured to produce the modified sugar products,
CC   e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
CC   tetracycline, polyene, polyether, ansamycin or isochromanequinone.
CC   The present sequence is Streptomyces venezuelae sugar (desosamine)
CC   biosynthetic gene cluster.
XX
```

SQ    Sequence 13613 BP; 1858 A; 4732 C; 5093 G; 1930 T; 0 other;

Query Match    7.7%;    Score 527.4;    DB 24;    Length 13613;
Best Local Similarity    53.3%;    Pred. No. 4e-57;
Matches 1475;    Conservative    0;    Mismatches 1181;    Indels 109;    Gaps 13;

```
Qy    2269 TGAAGGCCCTGGTCCTTGGCAGTGAACCGGCAGCAGACTGAGGCGGTCACCCACACCG 2328
           |||| ||||    | |||  |||||| ||  | |  | |||| |||| |   |||| ||
Db    9988 TGAAGGGAATAGTCCTTGGCGGGAGCGGAGCTCGGCTGCATCCGGCGACTCGGGTCA 9929

Qy    2329 CCGCCAAGCAGCTGCTCCCCATCGCCACAAGCCCGTGCTTCTTCTACGCCGTGGAGTCCC 2388
           ||  ||     | |||  | ||||  |  ||||||||| |||| |||| ||||| |||||
Db    9928 TTTCGGAGCCAGATTCTTCCGGTCACAACCGATGATCTACTAGTCGTCGTCGTGGGTTC 9869

Qy    2389 TCGGCCGGCGGGGTGTCCGGAGGCCGGAGATTCAAATCATCTCGGGTTACGGCGGAGATCT 2448
           | | |||||||  | | ||   || |||  |  ||| ||| | ||| ||||| ||||| |
Db    9868 TCATGCTCGGCGGTATTCGCGGAGATTCAAATCATCTCCAGCACATGACGCCTTCGTCCA 9809

Qy    2449 GCGA--ACTCACCGGCGACGCACCGTCGGTTCGGGTTACGGTCCAGCATCAACTACTCCACCAGC 2505
           | |||   ||| ||||||||||||||||||||  |||||  | |||| |||||  ||||||||||
Db    9808 TCCAGTGCGTTCTCGGAAACGGCAGGGCACCTGGAATAGAACTCGACTATGCGGTCCAGA 9749

Qy    2506 CCCGCCCGCTCGGTCTCGCGCACCGCGTTGCCGACGCACGCACTTCTGTGACGAGCACG 2565
           |||| ||| ||||||  ||||  |  |||||||  || |||||| ||||| |||||||||
Db    9748 AAGAGCCCGAGGAATCGCACCGCACTTCTGTCGCGACGCACATCGGCGACGACGACA 9689

Qy    2566 ACTTCCTGCTGTACCTGGGGGACAAACTACCTGCCC---CAGGGCGTCACCGGACTTCGCCCC 2622
           ||||| ||  | ||| || |||||||  | | ||||   ||||||||||||||| ||||||||
Db    9688 CCTGCGCCCTGATCCTGGGGCGACAACATCTTCCACGGGCCCCGGCTCTCACCGCGGCTCTGC 9629

Qy    2623 GCCAATCGCCGCGAGTCCCGCGCCGATCCCGCCGCTCGTGCCCGGCTGCGGGCCCGGACCCGT 2682
           | | || |||| ||| ||||| |  |  ||  |||| ||| |||||||  ||| |||||||||
Db    9628 GGGACAGCATCGCGCGTCGCGAGGTGCGGGACGCGGAACGGCTGCGGTTGACCGGAGAAAC 9569

Qy    2683 CCGGCTTCGGCGGTGCGGAGCGTGGGCGAGGTGGACGCGACGCACTCGTTCAGCCGGAAGC 2742
           |  |||| |||||| || |||| ||||||||||||||||||||  ||||||||||| ||||
Db    9568 AGCGGGTACGGCGGGTACGGCGAGCTGCTGCTCAGCCGTGTACGCCGTGTCCCGGCGTC 9509

Qy    2743 CCGACGTCCCGGCGCAGCTGCGCCTCGCCGGCTCTTACCGTCACCCTCCGCCGCCGCGCGGGCGAGCTGGAGATCACCCGACCG 2802
           ||| ||||||||| |||  | |  | || |  ||| | ||||  ||||| |||| || |||||| ||||| |  ||||||||
Db    9508 CCGTCAAGCCGCGCTCCAACTCGGCCGCTACCCTCTCTACCGTCTACGGAGCTGGAGATCTGGAGAGTCGCGCCCCGAGCGTCA 9449

Qy    2803 ACGAGGCGGTACGGGCCATCACCCCCTCGCGCGCGGGCCGCGCGGGCGAGCTGCGGCCGGAGATCGGGCACCACCGGAGATCA 2862
           |||||| |||| |  |   |||| || |||||||| ||||||||||| | ||||| |||| |  ||||| |||||||| |||
Db    9448 TCGACATCGCCAAGAACACCGCCCTCGCCGGGGCGGGCTCGGCCGGGCTGCGCGGGGCGAGCTCGACGTCA 9389

Qy    2863 TGCAGTGGATGATCGACCGGGGCCTTGCGCTTGGCGGCCGGA--GACCCACCACCGCCCGGCTTCGGCT 2919
           ||||| | |  |||||||| |  |  | ||||  ||||  ||  ||||| || ||||||  |  |||||
Db    9388 ACCGGCGTCTACCTGGACGCGGCGGGCGGCCCCGAGCTCGTCAACCTGGTCAACGCCGTCAGCTGGCCT 9329

Qy    2920 GGGCGGGGACACCGCAGCGCGGAGAGCAGACATGCTGCCTTCGCTCGCTGGGCGGGGCGACCGGAC 2979
           |||| ||||||||| ||||||||||| |||||||| | |   | | | ||||||||| | ||||
Db    9328 GGGCTGGACACACCGGAGGTCGGCGCACACTCGCTGGAGGTCGTGCGGCGGCGGCCCCAGTACGGTCCTGG 9269

Qy    2980 TGGAGGGCCGGCAATCGGAGGGGGAAGGTCGACGAGCAGCAGTCGACACGGCTGGCGTGGGCCGACCTGTGGGCG 3039
           ||||||||| || ||  |||||| |  | | ||||||| | ||| |||||||||   ||  ||||| ||||||||
Db    9268 AGGAGCG------------GCAGGGCGGGTCGTCGGATCGCGGCGTCCCCGGGGGCGT 3219

Qy    3040 TGGCCGAAGGCGGCGATCGCCTTCGTGCGGTGCTGGCTGCCCTGCACGTGGTGGGCGGCCGG 3099
           | |||| || |||||||||||  || | | | |  |  |  ||| |||| |||||||| |||
Db    9241 GCCTTGAGGAGTCTCAGCAACTTCCCGCGATCGCCTTCATCGGGCCGGGAGGACTCGGGG 3159

Qy    3100 GTGCCGGTCGCTCAGCGACTCCCAGTGTCCAGTACACGTCGTCGGGGCGGAGTCCGGGGC 3159
           ||||| ||||  |||| ||  |  | || ||  | ||||   | |||  | | | | |||
Db    9181 GAGAAGGCCTCTCCTCCGGGCCATCGGCCATCCGGAGATCCGCGGCT 3129

Qy    3160 TCGGAGACAGGCCATCGATCAGTACTCCGTCCGTTCCTGCCGGGCCCCGGGCCGACCACCGACAGCAGCGGCGT 3219
           |  || ||   | ||||   ||||| |||||| || || ||||| ||||||| | || ||  |||||||| |
Db    9128 GCTCGCCGCGACAATGACGTCGG--------GCCGGACCGGCAACCGGGCGTCCGGAGGTGTCCGGCGAGCAGCGGC 9069

Qy    3220 CCGGTCATCGAGGCGGGCGCGGAGAGGTGTCCCTTCATCGCGCCGGCGCGGCCGCGCCCCCGTCTCC 3279
           ||   |||  ||| |  | ||| ||| |  |  | |  |||| |||| |||||||||||||||||
Db    9068 CCGACAGTGCGACCCGCCACCGCCGCCGCCGCCGCAACCCCGACAGTGCGACC------------- 9019
```

```
Qy  4345 CACACCGCCGCCGCGGCCGGTGCCGCCGGTCAGCGTTCCGTGAGCGGCCGGCGACCGGCCG 4404
              |  ||||||| |||||||    |  | ||| | ||   |     | | |||  ||
Db  7957 GCCGAGACCCCCGCGTCCCCTTCCTGACCCCGGACCAC                        7898

Qy  4405 CCCCGGGCGGGGCCGGCCGGTGGACCCCCGGACCAC---CAGTTCCGGCGGCATGAAGAGCGAA 4460
              |   ||||  |||  | ||||    ||||  ||||   |||| | |   || ||| ||||
Db  7897 GAGACCGACGCGCCGATCGCCCGCTCCGACTCGGGGCGCTACCTCCTCGACCCGAA        7838

Qy  4461 TTCGGTGCCGGCGGCGGCGTTCCGCTCATCTCCTCAGCAGTGCGTCACGGCGACCTG 4520
              | | || | |  ||| |||| | || |  |||| ||| | ||| ||   |||
Db  7837 CTCGAAGGATTCGAGGCGGCGAGTTCGCCGGCTACTGCGGCGGACGCCGTCGGCGTG 7778

Qy  4521 CCCCATCGCCTTGACGGGCT-------GTCTGATGGTGGTCAGGGGGACGGTCGGTGAAG 4572
              | || ||| |||| |||||       |    ||  | || |   ||| | | |||
Db  7777 AACAGCGGGATGGACGACGCCCTCCAGCTCCGCCCTCCGGCCTCGACCCCGGGGAC        7718

Qy  4573 GCCATGAGCGGCGGAGTCGTCGAAGCGACCACCGAGATGTCACCGGGAACCGTGAGACCC 4632
              ||| || || ||| ||| ||  | |  |  |||| |  |  | | |  ||| | ||
Db  7717 GAGCTGATCGTCCCCTCCGCACACGTACATCGCCAGCTGGCTCGCCGTGTTCCGCACCGGC 7658

Qy  4633 CGGCCGGCGCGGCGGCCCGCACGGGCGGCCCGCCCATCATGTCGCTGGCGCACATGACGGCG 4692
              | ||||  ||| |||     |||| ||  |   | |||  | |  || |  || ||  |
Db  7657 GCGACCCCGTGCCCCGTGCAGCCGCACGAGGACCACCCCACCACCCGGACCCGCTGCTCTGTC 7598

Qy  4693 GTGCAGGCCCAGGTCGATCAGCGCGGACGCGCG-CGGCCTGGCCCCCTCCAGGGAGGAAACAG 4751
              |||| | | | || |  |||||  |  || || |     | || | |  | ||||| | ||
Db  7597 GAGAAGGCGGATCACCCCCGGCACCCGGGCGTCCTCCCGGCCTCTCACCCTCTACGGGCACCCC 7538

Qy  4752 CGAGTGCTGCACGAGCTCCTCGGACTCCCGCGGCCGACACTCCCCAGGTGCTCCCCCACGCC 4811
              || ||| | || |||| | || || ||| |  |  ||  | |  |  |  | | |||| ||
Db  7537 GCCACATGGACGCGCCCTCCGCGGCCTCGACCGCGCACATCGTCGAGGAC               7478

Qy  4812 GGGCCGGAACCCCTCGATCTTCCGCTGCACCGGCACGGGGGCCGGCCGACGCGGCGAG 4871
              |||| | | | |  | | || || | | |||| | |  ||  || | |||  | |
Db  7477 GCGCCGCAGGCCCACG--GCGCCCGCTACCGGGGCCGCGGATCGGCGCCGGGCGTCGTCG 7421

Qy  4872 GCCGCCATCCGCGTGCCCCCAGCTCCGCCAGGTGCGCCACGGCGCATCGCCGCCCG 4931
              || |||  |||| |    | || | | ||   || | || || ||| |  |||
Db  7420 GTGGCCGCGTTCAGCTTCTCACCGGCCAAGAACCTCGGCCGTGCTTCGGCGACGGCCGCGCC 7361

Qy  4932 GTCGTCCGGGAGACGAAGGGTGCCTCCGATCCGGGGCGGAGAAACCCGTTCACGAGGACGAA 4991
              |  | | ||| ||||  ||  | ||  | | |||| ||| | || |  | ||| ||
Db  7360 GTCGTCACCGGCGACCCCGAGCTCGCCGAACGGGCTCCGATGCGTCCGCAACTACGGCTCG 7301

Qy  4992 GGGCA                                                      4996
              ||||
Db  7300 CGGCA                                                      7296
```

RESULT 11
AAT76906
ID   AAT76906 standard; DNA; 546 BP.
XX
AC   AAT76906;
XX
DT   02-JUL-1998  (first entry)
XX
DE   S. glaucescens acbD DNA fragment.
XX
KW   Acarbose biosynthesis; acbA gene; acbB gene; acbC gene; acbD gene;
KW   acbE gene; acbF gene; enzyme; alpha-amylase inhibitor; treatment;
KW   diabetes; ss.
XX
OS   Streptomyces glaucescens.
XX
PN   DE19622783-A1.
XX
PD   11-DEC-1997.
XX
PF   07-JUN-1996;  96DE-1022783.
XX
PR   07-JUN-1996;  96DE-1022783.
XX

PA   (FARH ) HOECHST AG.
XX
PI   Decker H;
XX
DR   WPI; 1998-033827/04.
XX
PT   Recombinant DNA molecule comprising genes for biosynthesis of
PT   acarbose - an alpha-amylase inhibitor useful in treatment of
PT   diabetes
XX
PS   Example 2; Page 7; 35pp; German.
XX
CC   This sequence encodes a fragment of the Streptomyces glaucescens GLA.O
CC   acbD gene which has been generated via PCR. The acbD gene is involved in
CC   the acarbose biosynthesis pathway. This gene and other acarbose
CC   biosynthesis genes such as those represented in AAT76903 are useful for
CC   producing acarbose, which is an alpha-amylase inhibitor useful in the
CC   treatment of diabetes.
XX
SQ   Sequence 546 BP; 69 A; 235 C; 168 G; 74 T; 0 other;

Query Match             7.6%;  Score 522.6;  DB 19;  Length 546;
Best Local Similarity   97.4%;
Matches  531;  Conservative   0;  Mismatches  14;  Indels   0;  Gaps   0;

```
Qy  3352 CACCGGCGGAGCGGGCTTCATTCGCGGGCTTCCGGCTACGTCCGCCCACGTCCGTCCGCCCGGGGC 3411
          |  | ||| || || || ||| | |||| | |||| |||||||||| || |||||| | |||||||
Db     1 -CCCGGGCGGGGCGGGTTCATCGGCGTCCGCCTACGTCCGCCTCCGCCTCCGTCGCCCGGGGC    60

Qy  3412 CCCCGGGCGGGCCGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCCCGCAGCCTCGCCCG 3471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CCCCGGGCGGGCCGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCCCGCAGCCTCGCCCCG 120

Qy  3472 CCTGCACGCGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGGACGTGTGCGACAC 3531
          |||||||||| ||||||||||||||||||||||||||||||||| ||||||||||||
Db   121 CCTGCACGCGGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACAC 180

Qy  3532 CGGCGTCGTCGACACGCGTGGCCCGCGCCACGACGACGCATCGTCTTCGGCGCCGAGTC 3591
          || |||||||||||| | |||||||||||||||||||||||||||||||||||||||||
Db   181 CGCGCTCGTCGACACGCTGGCCCGCGCCACGACGACGCATCGTCTTCGGCGCCGAGTC 240

Qy  3592 GCAGTCGACCGTCCATCACCGACAGCGGTGCCTTCACCCGCACCAACGTGTGGGCAC 3651
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCACGTCGACCGTCCATCACCGACAGCGGTGCCTTCACCCGCACCAACGTGTGGGCAC 300

Qy  3652 CCAGGTCCTGCTCGACGCGCCGGCTCCGACGCGTGTGGCCACGGTGCGCACGTCTCCAC 3711
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCAGGTCCTGCTCGACGCGCCGGCTCCGACGCGTGTGGCCACGGTGCGCACGTCTCCAC 360

Qy  3712 CGACGAGGTGTACGGCGTCCCTCCGCACGGGCGGCCGCCGCGGGAGCGGACCCCCGCTTCC 3771
          |||||||||| ||||| || || |||||||||||||||||||||||||||||||||||||||
Db   361 CGACGAGGTGTACGGCGTCCCTCCGCACGGGCGGCCGCCGCGGGAGCGGACCCCCGCTCCC 420

Qy  3772 GACCTCGCCGTACGGCGCGTCGAAGGCGCCGCCTCGGACCTCCATGGGCGCTCGCCCACCACCG 3831
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACCTCGCCGTACGGCCGCGTCGAAGGCGCCGCCTCGGACCTCCATGGGCGCTCGCCCACCACCG 480

Qy  3832 CACCCACGGCCTGGACGTCCGGGTGACCGTGTTCGAACAACTTCGGCCCCCCACCAGCA 3891
          |||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||
Db   481 CACCCACCGGCCTGGACGTCCGGGTGACCGTGTTCGAACAACTACGCCCGCACCAGTT 540

Qy  3892 TCCCG 3896
          |||
Db   541 CCCGG 545
```

RESULT 12
AAQ39093
ID   AAQ39093 standard; DNA; 2634 BP.
XX
AC   AAQ39093;
XX
DT   03-AUG-1993  (first entry)

```
XX
DE   Streptomyces nodosus 2634bp BamHI fragment.
XX
KW   snoT; snoD; snoM: microbial synthesis; actinomyces; hybrid;
KW   glycosylated; natural products; prods.; sequencing; gene;
KW   polymerase chain reaction; secondary metabolite biosynthesis; ss.
XX
OS   Streptomyces nodosus DSM40109.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..404
FT                  /*tag=  a
FT                  /note=  "snoM C-terminal"
FT   CDS            416..1534
FT                  /*tag=  b
FT                  /note=  "snoT sequence"
FT   CDS            1561..2628
FT                  /*tag=  c
FT                  /note=  "snoD sequence"
XX
PN   WO9306219-A.
XX
PD   01-APR-1993.
XX
PF   15-SEP-1992;   92WO-EP02111.
XX
PR   18-SEP-1991;   91DE-4130967.
XX
PA   (FARH ) HOECHST AG.
XX
PI   Brau B, Distler J, Grabley S, Piepersberg W, Sichel P;
PI   Stockmann M, Taleghani KM;
XX
DR   WPI; 1993-117540/14.
DR   P-PSDB; AAR34012, AAR38296, AAR38297.
XX
PT   Sec. metabolite biosynthesis genes from Actinomycetes - isolatable
PT   with hybridisation probes using DNA, useful in microbial synthesis
PT   of glycosylated and natural prods. in Actinomyces
XX
PS   Claim 1; Fig 4; 38pp; German.
XX
CC   The sequence is that of a 2634 BamHI fragment from Streptomyces
CC   nodosus which comprises the complete snoT sequence (encoding
CC   amphotheronolide B-dTDP-D- mycosaminyl transferase), the snoD
CC   sequence (encodes dTDP-D-glucose synthase) and the partial snoM
CC   sequence (encoding dTDP-4-keto-6-deoxy-D-glucose isomerase). These
CC   isolated genes can be used for, the microbial synthesis of hybrid
CC   and glycosylated naturalk products in Actinomyces; increasing the
CC   secondary metabolite yield in Actinomyces; the isolation of enzymes
CC   in biosynthesis and their use in enzymatic synthesis; the glycosylation
CC   of cpds. by biotransformation in Actinomyces; the identification of
CC   structurally and functionally novel genes; and the screening of
CC   producers of secondary metabolites.
XX
SQ   Sequence 2634 BP; 392 A; 981 C; 892 G; 369 T; 0 other;

Query Match     7.3%;   Score 497.6; DB 14;   Length 2634;
Best Local Similarity  66.3%;   Pred. No. 2.4e-53;
Matches  716;   Conservive   0;   Mismatches  364;   Indels    0;  Gaps    0;
```

```
Qy  2256 GCGGAGGCGGCGAGGTGAAGGCCCTGGCTCCTGGCCAGGTGAACCGGCCAGCAGACTGAGGCCG 2315
          || |||   |  |  |||||||| ||||| || ||| |||| | ||||| |  ||| ||||||
Db  1549 GAGGACCTTCGATGAAGGCTCTGGTGCTCGCCGGGGATCTGGTACCGGCCCTGCGGCCT 1608

Qy  2316 TTCACCCACACCGCCCCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTTCTTTAC 2375
          ||||   |  ||    |  |||    |||| | |||||||||  ||||| ||||| |  ||
Db  1609 TTCAGTTATTCGATGCCAAACAACTGATCCCCAACGCCAACACACCCCGTGCTTGGTGCAT 1668

Qy  2376 GCGCTGGAGTCCCTCGCCGGGGTGTCCGGGAGGCGGGGCCGTCGTGTGGGGCGGTAC 2435
          ||||||| || |  | || | |  |||||| |||||   |||||   || ||
Db  1669 GTGCTGAACGCCGTCCGGGAGCTGGGCGTGACCGAGGTCGGCGTCATCGTCGGCAACCGC 1728
```

```
RESULT 13
AAA39283/c
ID   AAA39283 standard; DNA; 16020 BP.
XX
AC   AAA39283;
XX
DT   08-SEP-2000  (first entry)
XX
DE   Streptomyces nogalater nogalamycin biosynthesis gene cluster SEQ ID NO:1.
XX
KW   Streptomyces nogalater; nogalamycin biosynthesis; antibiotic;
KW   anthracycline biosynthetic pathway; gene cluster; drug screening;
```

```
Qy  2436 GGCCGGGGAGATCCCGACAACTCACCCGCACGGCACCGGCCGCTTCGGGGTTACGCATCACCTAC 2495
          |||| ||||| |  | || ||||||  |  ||||||||||||| ||||| |||||||||||||||
Db  1729 GGCCCGAGATCGAGGCCGTTGCTCGGCGGTGCCCGGTTCGACGTGGCATCACCTAC 1788

Qy  2496 CTCCACCAGCCCGCCCTCGGTCTCGGCCACGGTGCTCGCCGGTCGGTGGCCCCCGGCCTTCCTG 2555
          |||| |  ||||||||||| | | |||| ||||  |||||||| |  | |||||||||||||||
Db  1789 ATCCCCCAGGACGCACCGGCTGGCCCACACCGTSTCCATCGCCCCGGCCTTCCTC 1848

Qy  2556 GGCGACGACGACTTCCTGCTACCTGGGGACAACTACCTGCCCCAGGGCGTCACCGAC 2615
          ||||||||||||||| ||| || |||||| | |||||| |||||| |||||| |||||
Db  1849 GGCGACGACGACTTCGTGATGTACCTCGGCGACAACATGCTGCCCGACGGAGTCACCGAG 1908

Qy  2616 TTCGCCCCCAATCGGCCCGCCGATCCGGCGGCCCGGCTGCTCCTCACCCCGGTCGCG 2675
          |||||  | ||  |||   ||||||||| |  ||||  || | | |   || ||||||
Db  1909 ATCGCCGAGGAGTCACCCGCAGCGCGCCGCCCAGGTCGTCGTGCACAAGGTCCCC 1968

Qy  2676 GACCCGTCCGCCTTCGGCGTCGGAGGTGCGACGCGGACGGGAAGCTGCGTGCGCCTGGAG 2735
          |||| ||||| |||||||||||| | | ||||| ||| |||||||||||| |||||||||
Db  1969 GACCCGCGCTCCTTCGGCGTCGCCGAACTCGGCCCGCGCCGACGGGAGGTGCTCGCCCTGGTG 2028

Qy  2736 GAGAAACCCGACGTCCCGCAGCTCGCTCGGCCCAGCTCGTTCCGCCCGGCCTGTACGGCCTCAGCCCG 2795
          ||||| ||       |||||||  || ||||| ||||||| || |||| ||||||||||||||||
Db  2029 GAGAGCCCGTGGCAGCCGCAGCGACATGGCCCGATCGGGGTCTACTTCTTCACCGCC 2088

Qy  2796 GCCGTCCACGAGGCGGTACGGGCCCATCACCCCGCTCCGCCCGGAGCTGGAGATCACC 2855
          || |||  ||||||| | |||  ||||| |||| || |||||||| ||||  |||||||
Db  2089 GCCATCCACCAGGCCGGTGGCCGGCCATCTCGCCCAGCAGCGCCGAACTGAGATCACC 2148

Qy  2856 CACGCCGTGCAGTGGATGATGATCGACCGGGGCCTCGCGGGCCTGCGGGCCGAGACCACACCCGG 2915
          ||||| ||| |||| ||  | | || | |||||  ||||  ||||  |  |  ||||||||||||||
Db  2149 GACGCCGTCCAGTGGTTGGTCACCTCCGGCGACGTGCGCGACGAGTGCGGCCAGCCTCTACGACGGC 2208

Qy  2916 CCCTGGCCGACACCGGCCAGCCGCCGGAGGACATGCTGGGAGGTCAACCGTCAACGTCCTGGAC 2975
          ||    | ||||   || ||||| |  |  |  ||||  ||||  || |||   |||||||||||
Db  2209 TACTGGAAGGACACCGGACGGGGTCGAGGACGTCCTTGAGTGCAACAGCCCACCTCCTGAC 2268

Qy  2976 GGACTGGAGGGCCGCATCGAGGGGAAGGTGCACGCGCACAGCACGCTGGTTCGGCGGGGTC 3035
          ||    |||  || ||||| |  |||||| ||||| ||||| |  |||||| |  ||||||
Db  2269 GGCCTGACCCCGCGGCTGCGACGGACAGGTCGACGCGGTGCTCGTCGGTCGGGGACTGC 2328

Qy  3036 CGGGTGGCCGAAGGCGCGATCGTGCGGGGTCACACGTGGTGGGCCCGGTGGTCATCGGC 3095
          | ||  |||  || |||| || ||| | |   || || |  ||  ||  ||||||||||
Db  2329 GTGATCGAGGCCGGGGGCGGCCCGATCGTGCGGTCGCGGGTCGAGGGCCCGCGATCATCGGC 2388

Qy  3096 GTGGTGCCGTCGTCAGCAACTCCAGTGTGCGGCCGTACACCTCCATCGGGGAGGACTGC 3155
          ||||| |  ||   |||||||| |  | |||  ||  || |||| || | ||||||||||
Db  2389 GCGGGCACGGTCCTTCAGGACAGCAGCCAGGTGGCCCGCACCTCCATCGGGCGGCGGGACTGC 2448

Qy  3156 CGGGTCGAGGAGCAGCCGCCATCGAGCACCGCCGAGTACTCCG 3215
          |  ||||| || ||||  || | |||||||| |  | ||| ||||||
Db  2449 ACGGTGACGGACAGCCGGCTGGAGGGGCTCCATCGCCCTGGACGGGGTCGGGCTCACCGGC 2508

Qy  3216 GCGTCCCCATCGAGGGCGTCCCTCATCGGCCGCCGCCGCTCGTCGGCCGCCCCCCGT 3275
          |  | |||| || | |||||| |||| ||   | | || || ||   ||| |||| |
Db  2509 GTGCCCGGCGCTGCCGCAACTCGCTCATCGGCCGCTCGCCGTCCGGCGACCACCGCGCCCCC 2568

Qy  3276 CTCCCGCCAGGCTCACCGACTGGTGATCCGGACGCAAGGTGTATCTCACCCCATGA 3335
          ||| | || |  |  ||||||||||| |||| |  || ||| |  ||| |  ||||
Db  2569 GGCACGGGCCATCACTGCCTGGTCGTCGGAGAGCACCACCCCGAGTGGAGGTCGGGCATGA 2628
```

KW   antibiotic; antitumour antibiotic; anthracyclinone; ds.
XX
OS   Streptomyces nogalater.
XX
PN   WO20024775-A1.
XX
PD   04-MAY-2000.
XX
PF   20-OCT-1999;   99WO-FI00870.
XX
PR   23-OCT-1998;   98FI-0002295.
XX
PA   (GALI-) GALILAEUS OY.
XX
PI   Ylihonko K,  Torkkel S,  Palmu K,  Hakala J;
XX
DR   WPI: 2000-350696/30.
DR   P-PSDB; AAY91055, AAY91056, AAY91057, AAY91058, AAY91059, AAY91060,
DR   AAY91061, AAY91062, AAY91063, AAY91064, AAY91065, AAY91066, AAY91067,
DR   AAY91068, AAY91069, AAY91070, AAY91071.
XX
PT   Isolated and purified DNA fragment for obtaining novel hybrid
PT   antibiotics comprises the gene cluster for the anthracycline
PT   biosynthetic pathway of the bacterium Streptomyces nogalater -
XX
PS   Claim 2; Page 32-40; 59pp; English.
XX
CC   The present sequence represents the nogalamycin biosynthesis gene
CC   cluster isolated from Streptomyces nogalater. Nogalamycin is an
CC   anthracycline antibiotic, so the nogalamycin biosynthetic pathway is
CC   also known as the anthracycline biosynthetic pathway. DNA fragments,
CC   plasmids and process from the present invention are useful for obtaining
CC   novel hybrid antibiotics, such as anthracyclines (antitumour antibiotics)
CC   or anthracyclinones.
XX
SQ   Sequence 16020 BP; 2295 A; 6011 C; 5425 G; 2189 T; 100 other;

Query Match             7.1%;  Score 486.8;  DB 21;  Length 16020;
Best Local Similarity  61.7%;  Pred. No. 3.8e-52;
Matches  910;  Conservative    0;  Mismatches  532;  Indels   34;  Gaps    7;

Qy   3197 CGGCGCCCAGGTCGAGGGGGCGTCCCGCATCGAGGCGTCCCTCATCGGCCGCCGGCGCCGGA 3256
          || ||| ||||| ||||||| ||| || ||  || |  | ||| | ||| || ||||| | ||
Db   8410 CGGCGGTCGAGCCGCGCTCAACGCGTCAACCGCTCCTCCACGGCCGCCCCCGACGC 8351

Qy   3257 CGTCGGCGCCCCCGGTCTCCCGGCGCCCCCGGGCGCGGCGGCCGCCGGTGGCGGTCAGC 3316
          ||| |||||| |   | |    |  | | || ||||   ||| | || || |||||
Db   8350 TGGGTCGAGGGGGCCGCACTCCCGCGCCCCGACGCACCCATGGCGT 8291

Qy   3317 GGTGTATCCACC--CCATGACCACGACCATCCTCGTCACGGCGGGCGGCTTCATTC 3374
          | | | | |    | | || ||||   || ||| |  || |||| |  || |
Db   8290 CGCACACCTCCGCGACGACGGTGAACATCCTGGTCACGGCCGCGCCGGGGGCCGTCGGG 8231

Qy   3375 GCTCCCGCTACGTCCCGGCCGTCTGCGTGTCGGGGGCCCCGGGGGCCCGGTCGCG------- 3427
          ||||| ||  | | |||   | | ||| | |||| |||  | ||| |  || |
Db   3427

Qy   3375 GCTCCCGCTACGTCCCGGCCGTCTGCGTGTCGGGGGCCCCGGGGGCCCGGTCGCG 3427
          |||||| |||||| | |||  | | ||| ||||| |||| || ||| |  || |
Db   8230 GCTCGGCCGTATGTCGCCATGTCGGGAGAACCGGCGCCGGGGGGGCCCCGGCGG 8171

Qy   3428 -----GTGACCGTCCTGACAAACTCACCTACGGCCGCAGCCTGCCCCGCACGCG 3482
              |  | |||| || |||| |||| |||| ||  | | |||| ||| |||||
Db   8170 TACGCCGTTCGGCCGTGCGTTCGCGCCGCACATCCTCGACGTCCGCGGCGGTGGTCG 8111

Qy   3483 TGCGTGACCATCCCGGCCCTCGTCGGACCGGCTGCGTGCGACACCGGCCGCTCGTCG 3542
          | ||||| ||| | ||| |||| | ||||  ||| ||||   | || ||| |||||
Db   8110 TACG---CGGTGACCGGCGCCGCGGCGGCACACGGCCCGGAGCTGCGCGCCGACGT 8054

Qy   3543 ACACGCTGGCCGCGCGGCACGACGACATCGTGACCGGCCTCGTCGACC 3602
          | |||||||| ||| |    | ||||  | | ||| ||| || || ||
Db   8053 ACGAGCTGATGGCACACAGCGACCACTTCGCGCCACTTCGCGCCAGCCGC 7994

Qy   3603 GCTCCATCACCGACGAGCGTGCCTTCACCCGACCAAACGTGCTCTGGCCACCCAGGTCCTGC 3662
          ||  | |||| ||  || | ||||| | |||  | | |||| |  ||||| ||  |||||||
Db   7993 GGTCCATCCGCGCCGACGACTTCGTCGCGCTGCTGAACGTCGTCAAGCGTCGCCGCCAGGCGCTGC 7934

Qy   3663 TCGACCCGCCGCTCCGCCACGGTGTCGCGCACCTTCGTGCGCACGTCTCCACCGACGAGGTGT 3722
          |||||| ||| ||| | ||||||| ||||||||||| |  | || |  || ||||| ||||||
Db   7933 TGGACGCGGCCGCTGCGCACGGGGGCGCGCACCCTTCGTTGTCCTGTGTCACGGACGAGTCT 7874

Qy   3723 ACGGGGCTCCTCCCGCACGGGGCGCGGGGCCCGGAGAGGACGACCCCTTCGTTCCGACCTCGCCGT 3782
          |||| || |  ||||| |||| || ||  ||| || ||||||| | |  || | |||||| ||||
Db   7873 ACGGCTCCATGCCTCGGTCCTGCGCCGGTCAGCCGTCAACCTGCCGACCTCGCCCCAACTCCCCGT 7814

Qy   3783 ACGGCGGGCTCGAAGGCGGCCTCGAGACTCATGGCGCTGACCTCATCGCCTCCACCGCACCCAGGCC 3842
          |||| || ||||| ||| |  | || || ||  ||  |||| || ||||| ||||| ||| |||||
Db   7813 ACGCCGCCTCCAAGGCGTCGGCCGACCTGGCCGACCTCGATGGCGTTCGCTGCCGACCGCACCGCGCC 7754

Qy   3843 TGGACGTCCGGTGACCGGCTGTTCGAACAACTTCGGCCCCCACCAGCATCCCGAGAAGC 3902
          |||||| ||||| || ||||| || | |  || | ||  | | |||||  | || |||
Db   7753 TGGACGTGCGGGTGCGCCGTGTCCCACAACTACACGGCGTCAGCATCCCGCTCAGCGCGAGAAGC 7694

Qy   3903 TCATACCGGCTTCCTGACCAGCCTCCTGTCCGTCGGACGTCGGGCACCGTTCCCCTCTACGGGCACG 3962
          | |  |||||||| || ||  ||  | || ||||| || | || || |||| || | ||||| ||
Db   7693 TCATCCCGGCGTTCGTCACCAACTGTTGGACGGGGTGTCGCGCTGCCGTGTACGGGGACG 7634

Qy   3963 GGCGGCACGTGCGGCGACTGGCCTGCACGTCGACGACCACGTCAGGGCCGTCGAACTCGTCC 4022
          | |   | ||| | ||  | ||||  ||  |||| |||   | || ||| | ||||| | ||
Db   7633 GGCGGAACGTGCGGCGGAGTGCGGCCGGTCAGTTGGAGGAGGACCACTGGCGCTGGACCTGGTGC 7574

Qy   4023 GCGTGTCGGGGCCCGGGGGCCGGGCCCGGAGAGATCTACACAATCGGGGGGCCGCACCTCGCTGCCCAACC 4082
          |  |||| ||| ||||||||| ||| || | | ||  | |||   ||||| || || |  |  | || |||
Db   7573 GCACCGGCGGGGGCCGGGGGCCGGCCGTGTGCCATCGGCTGTACCATCGCGCCGGGCTCCAACC 7514

Qy   4083 TGGAGCTCACGGTGGGTATGGTTCGTCCGCTGGGGCCGCCAGCCGCCGGCGCCGCCATCATCGTCCACG 4142
          ||||| | |||| |||  | ||| ||| ||||| |||   | ||| | |||  | | || | ||||||
Db   7513 TGGAGCTGGTGGGGTATGTTGCTGGAGCGCGTCGTTCCGGTCGTGCGGCGGTGGAGCGGTGGGGTCGAGCTCG 7454

Qy   4143 TCGAGAACCGCAAGGGCGCAGGACCGGCGCTACGCGGTCACCACAGCAAGATCACCGCGG 4202
          | || |||| |||| | |   || | ||  | | |  || ||||   | | ||| | |
Db   7453 TGCCCGACCGCAAGGGCACGACCACCGGCGTTACGAGCTTACAGCGGCGGCCAAGGGCCCGTGAGG 7394

Qy   4203 AACTCGGTTACCGGCCGCCGCCGACCGCACGGGGGCGACCACCCGCCGAAGTGGT 4262
          |||| | |  |||  | |  ||| | |   | ||||| | | ||||| || ||||
Db   7393 AACTGGCTACCGGCTACCGCGTCCTCCGGCGGAGTTCTCTCGCAGGTCGCAGTGGT 7334

Qy   4263 ACGAGGCCACGAGGACTGGTTGGCGTCCCCT-GCTCGCCGGCGACATGACGTGCGGCCGGA 4321
          ||  | || | | ||| |||  ||| ||||  || ||||   | ||  | || || ||  ||
Db   7333 ACCGGAGAACCGGTCCTGGTTGGGAGCCGGTCTCAAGCGGCGGCCCCCCGGCTGGCGGGCGGTA 7274

Qy   4322 CCGCAACCACCCGCCCGGCCACCGCCGACTGGTAGTGATCGGCGACCACAGCAA 4381
          ||| ||| |  | |||||| |||| | | | || |    || || |  | |||
Db   7273 CGTCCACCGTCGTCCCCGGCGTCGAGGCCGGTGAGCCGGAGATCGCGGCCGGTGAGCCGGAGAG- 7215

Qy   4382 GTCCGTGAGCCGGCCCCGGCCGTCTCACGGCGGGGCTTCATTC 4441
          | ||| | |||      ||  | |||   || | | | ||||
Db   7214 ------AAGGCATGCTCGCGCCGGCCCACCGGCGCTGGTTTCATCG 7161

Qy   4442 GTTCCGGCATGAAGACGAATTCGGCGGGGGCCCCGGCGGGCCCGGTCGCG------- 4501
          ||| || || | ||| ||  || || | || ||||  | ||| ||  ||
Db   7160 GGCCCGGCCGGGCGG---GTGTACTCGGCGCCATCGCCGGTCGGGGTTGGTGGTCAACTGCG 7104

Qy   4502 GTGCGTCCACGGCGACCTGCCCATCGCCTTCGACGGCGTCTGATGGTGTCAGGGGAG 4561
          || ||| || ||||||| || |  || || |||  | | | |||| | ||||||||
Db   7103 GGG------CGGTGGACGCGGCCCCAGTCGCGGCCGCACCGGCGGTGTGGTCAACTCG 7050

Qy   4562 GGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCGCGCACCGGCGCGGAA 4621
          ||||||| |  ||||  ||||| |  | | |||||| | | |||||||| |||
Db   7049 CCGCCTTCACGGACGTGACGGCGGCCGCGCCGCCGCGATGCGCGGTCAACG 6990

Qy   4622 CCGTGAGCCGGCCCCCGCGCCCCGCACCCGCGCGCGCGCGC 4657
          |  |||| |||||| | | |||  |||  |||| |||||||
Db   6989 GCGGCGGCCCCCGCGCTGCCCCAGCCGAGTCCCCGCGTGCGC 6954

RESULT 14
AAF88341
ID  AAF88341 standard; DNA; 990 BP.
XX

```
Qy 3578 TTCGGCGGCGAGTCGCAGTCGACGTCGACCGCTCCATCACCGACAGCGGTGCCTCACCCGCACC 3637
            |||||| ||||| |||||||||||||||||||||||| ||||||| ||||||
Db  247 TTCGCGGCCGGAAACCCACGTCGACCGCTCGATCACCGGCTTCGTGATCACC 306

Qy 3638 AACGTGCTGGGCACCCAGGTCCTGCTTCGACGCGCCGCTCCCGCACGGGCGCGGCCGAGATCGGCCACGTTC 3697
            |||||| | |||||| |||||| ||||| ||||||||||||||| ||||||
Db  307 AACGTGGTCGGCACCAACGTGCTGCTCCAGGCCGGCCGCTGTCGCCAGGCCGGGCCCGGCAAGTTC 366

Qy 3698 GTGCACGTCTCCACGGACGAGGTGTACGGCCTCCCTCCGCACGGGGCGCCGCGGGCCGGAGAGC 3757
            |||||||| |||||||||||| | |||||||   | ||  ||  | ||
Db  367 GTGCACGTTTCCACGGACGAGGTCTACGGCGTTCGCCTGGCCGGCCTCGTGGCCCGAAGAC 426

Qy 3758 GACCCCGTCGTTCGCCACCTCGCCGCGTCGACGGTCGAAGGGCGCTCGGACCTCATGGCG 3817
            || | ||||  ||||||| |||| ||||||| ||||| |||  || ||||||
Db  427 CACGCGCGTGGAGCCGGAATTCCCCGTACTCGGCGGCGAAAGCGGGCTCGGACCTCTGGCC 486

Qy 3818 CTCGGCCCACCACCGCACCCACGGCCTGGACGTCGCGGTGACGTCGCGGTTCGAACAACTTC 3877
            || ||| |||||||||||||||||| ||| | | |||||||||||  ||||||| |
Db  487 CGGCGCTACCACCGCACCCACGGACTGCCGGTGTGCATCACCCGGCTGCTCCAACACTAC 546

Qy 3878 GGCCCCCACCAGCATCCCAGAAGCTCATACCGGCCTTCCTGACCAGCCTCCTGTCCGGC 3937
            ||  | | || ||| ||||||||  | ||||||||   |  |  || |||||| ||||
Db  547 GGGCCCTACCAGTTCCCGGAGAAGTGCTCGCCGCTGTTCATCACGAACCTGATGGACGGC 606

Qy 3938 GGCACCGTTCCCCTCTACGGCGACGGGCGCACGTGCCGCAGCTGGCTGTCGACGTCGACGAC 3997
            || || ||||||| ||||||||||| ||||||| | |||||| | ||||||| ||||||
Db  607 AGCCAGGTGCCGCTCTACGGCGACGCGGCTCAACGTGGCGGCCTCAGCGAC 666

Qy 3998 CACGTCAGGGCCGTCGAACTCGTCCGCGTGTCGGGCGCCGGCGGAGATCTACAAACATC 4057
            |||||| ||| |||||||||||| | |  |||||||||||||||||||||| |||||
Db  667 CACTCGCCGGGCATCCAGCTGGTGCCGACTCCGGGCCGGGCCGGAGATCTACAAACATC 726

Qy 4058 GGGGGCGGCACCTGCCTGCCCAACCTGGAGCTCACCGACCACGGGTTGCTGCCACTGTGCGGC 4117
            |||||||||| |||||| ||||||||||| ||| ||||||| ||||||||||||| |||||
Db  727 GGCGGGCGGCCACCGACCGCTGACCAACAACGAGCTGACGAGCTGGCTGTCTGCGAGAGCTGGGC 786

Qy 4118 GCGGGCCGGAGCGCATCGTCCACGTCGAGAACCGCAAGGGCACGAGCGGCGCGTACGCG 4177
            ||||||| || ||||||||||||||||||||||||   |||||||||||||||| |||||
Db  787 CTCGACTGGTCGGTGGTGCGGGCCGGTCACCGGAACCCAAGGCCCAGGCCCACGACCCCGTACTCG 846

Qy 4178 GTCGACCACAGCAAGATCACCGGAACTCCGGTACGGCCCGCCACCGACTTCGCCGACC 4237
            ||||||||||||||||| |   ||||||||| ||  ||||||| ||||||||||||
Db  847 GTGGACCACCACAGCAAGATCGTCGAGGAACTGGGGTACGGCGCCGCAGTCGACTTCGAGAACC 906

Qy 4238 GCGGCTGGCCGACACCGCGAAGTGGTACGAGCGGCACGAGGACTGGTGGCGTCCCCTGCTC 4297
            | ||||||||| ||    |||||| |   | |  |||| |||| |||||| |    |||
Db  907 GGGCTGGCCGAGACAATCCGCTGGTACCAGGAGCAACCGGGACTGGTGGAGCCGCGCTGAAG 966

Qy 4298 GCC 4300
            |||
Db  967 GCC 969
```

AC   AAF88341;
XX
DT   28-AUG-2001 (first entry)
XX
DE   S. spinosa DNA fragment encoding ORF23, SEQ ID 52.
XX
KW   Forosamine: trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; dNTP-glucose-4,6-dehydratase; ds.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;     99DE-1057268.
XX
PR   27-AUG-1999;     99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   P-PSDB; AAB70970.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 7a; Page 336-337; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (1) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii), to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence, ORF 23, encodes an S. spinosa
CC   dNTP-glucose-4,6-dehydratase.
XX
SQ   Sequence 990 BP; 169 A; 338 C; 333 G; 150 T; 0 other;

Query Match     6.8%;  Score 465.8;  DB 22;  Length 990;
Best Local Similarity 68.5%;  Pred. No. 2.3e-49;
Matches 660;  Conservative 0;  Mismatches 297;  Indels 6;  Gaps 1;

```
Qy 3344 ATCCTCGTCACCGGCCGGAGCGGCGGCTTCATTCGCTCCGCCGCTACGTCCGCCGCTCCTGTCG 3403
            || |||||||||||| |||| || || ||||| | ||||  ||  |||  ||||||||
Db    7 ATTCTGGTCACCGGCGGAGCCGGAGCCGGTTTCATCGGCTCCACTACGTTCGGCAGTTGCTCGGT 66

Qy 3404 CCCGGGGCCCCCGG------CGGCGTCGCGGTGACCGTCGACAAACTCACCTACGCC 3457
            | |||  |||||||      ||||||||||||| ||||||||||| ||||||||||||
Db   67 GGTGCGGTACCCCGCCATTCGCCGACGCCGACGTGGTCGTGTCGACAAGCTCACCTACGCC 126

Qy 3458 GGCAGCCTGCCGCCCGCTGCACGCGGTGCGGTCGCGGTCGCGGACAACCCCGCCGGCTGAGGGC 3517
            ||||  | ||  ||  ||   ||| ||||||| |||||  ||||| |||||||||||
Db  127 GGCAACGAGGCGAACCTGGCCGGCCTGATGTCCGGCGGCCGGTGGTTGTCTGCGGGC 186

Qy 3518 GACGTGTCGACAACCGCCGCTCGTCGACACGTCGGCCGCGGCGGCGCGCACGACCATCGTGCAC 3577
            ||| ||||| ||  ||| ||  || ||  ||||||||||  ||||| || ||||| |||||
Db  187 GACATCTGGCACCCCGAACTGGTTGGCGGCCTGATGTCCGGCGGTTGCTCGGGTGTGCAC 246
```

RESULT 15
AAZ21502
ID   AAZ21502 standard; DNA; 2310 BP.
XX
AC   AAZ21502;
XX
DT   01-DEC-1999 (first entry)
XX
DE   Saccharopolyspora spinosa DNA fragment containing Rhamnose genes.
XX
KW   Rhamnose synthesis; glucose; glucose thymidylate; glucose dehydratase;
KW   epimerase; ketoreductase; gtt; gdh; epi; kre; insecticide; arachnid;
KW   spinosyn biosynthetic enzyme; insecticidal microlides; nematode; insect;
KW   polyketide; ss.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key            Location/Qualifiers
FT   CDS            88..1077
FT                  /*tag= a

FT                /product= gdh
FT                /note= "Glucose dehydratase"
FT    CDS         1165..1992
FT                /tag= b
FT                /product= kre
FT                /note= "Ketoreductase"
XX
PN    WO946387-A1.
XX
PD    16-SEP-1999.
XX
PF    16-FEB-1999;  99WO-US03212.
XX
PR    09-MAR-1998;  98US-0036987.
XX
PA    (DOWC ) DOW AGROSCIENCES LLC.
XX
PI    Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI    Treadway PJ,  Turner JR,  Waldron C;
XX
DR    WPI; 1999-551414/46.
DR    P-PSDB; AAY39320, AAY39321.
XX
PT    New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT    for production of insecticidal spinosyn compounds -
XX
PS    Claim 2; Page 172-175; 190pp; English.
XX
CC    This sequence is a DNA fragment of Saccharopolyspora spinosa, containing
CC    two genes involved in Rhamnose production. There are four enzymes
CC    involved in the production of rhamnose from glucose. The first is a
CC    glucose thymidylate transferase (gtt:AAY39322), the second is glucose
CC    dehydratase (gdh:AAY39320), the third is a an epimerase (epi:AAY39323)
CC    and the fourth a ketoreductase (kre:AAY39321). Kre and gdh genes are
CC    found in this sequence. These four enzymes are used during the
CC    biosynthesis of spinosyns and as a sole source of rhamnose for cell wall
CC    formation. Spinosyns are insecticidal microlides which are useful for the
CC    control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC    occurs via stepwise condensation and modification of carboxylic acid
CC    precursors generating a linear polyketide which is modified further. The
CC    DNA sequence AA221501 contains 24 genes and open reading frames involved
CC    in spinosyn biosynthesis, AAY39297-Y39319.  The genes identified in S.
CC    spinosa as having a role in the production of spinosyns are useful to
CC    improve yields of spinosyns, and for creating new spinosyns e.g. by
CC    mutagenesis, or interruption of steps in spinosyn biosynthesis. The
CC    modified spinosyns may be a new insect control agent or serve as
CC    substrates for further chemical modification and the creation of new
CC    semi-synthetic spinosyns. The genes are also useful to isolate similar
CC    sequences from S. spinosa or other species by hybridization.
XX
SQ    Sequence 2310 BP; 356 A; 759 C; 853 G; 342 T; 0 other;

Query Match             6.8%;  Score 465.8;  DB 20;  Length 2310;
Best Local Similarity   68.5%;
Matches  660;          Conservative    0;  Mismatches 297;  Indels   6;  Gaps     1;

Qy   3344  ATCCTCGTCACGGGCGGAGCGGGCTTCAATTCGGTCGCCGCCTACGTCGCCCGGCTCCTGTCG  3403
                |||| |||||||||||||||| || ||||| ||| || || ||  ||||  || || |||||
Db     94  ATTCTTGGTCACCGGCGGAGCCGGGTTTCAATCGGTCGCCCACTACGTTCGGCAGTTGCTCGGT   153

Qy   3404  CCCGGGGCCCCCGG------CGGCGTCGCCGTGACCGTCCTCGACAAACTCACCTACGCC  3457
                | |  ||||||   ||| |||||||||  ||||  |  || |||||||| |||||||||
Db    154  GGTCGCGTACCCCGCATTCGCCGCACCGCGACGTGGTCGTGCTCGACAAGCTCACCTACGCC   213

Qy   3458  GGCAGCCTCGCCACGCCGCTCGTCGCACGCCGTGCGTGACCATCCCCGGCCTCCACCTTCGTCCAGGGC  3517
                |||| ||| ||  |||| ||||||| ||||  ||| ||||  ||| |  ||| |||||  || |||||
Db    214  GGCAACGAGGCGAACCTGCCGCCGAACTGGTTGGCGCGCCGGACAACCCCGGCCTGAAGTTCGTCTGCGGC   273

Qy   3518  GACGTGTGCACACCGGCTCGTCGCACACGCCGGCCTCGACACATCGTGCAC  3577
                |||||||| ||||| ||| || |||| |  ||| |||| |||| | |||||
Db    274  GACATCTCGCACGGCACCGGAACTGGTTGCGCGGCCTGATGTCCGGACGTGGTGGTGTGCAC   333

Qy   3578  TTCGCGGGCCGAGTCGCACGTTCGACCGGCTCCATCACCGACGACGTCGCCTTCACCCGCACC  3637

```
                    GenCore version 5.1.4 p5_4578
              Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     May 28, 2003, 15:56:20 ; Search time 5709 Seconds
                                     (without alignments)
                                     19443.662 Million cell updates/sec

Title:      US-09-922-683-7
Perfect score: 6854
Sequence:   1 CTGCAGGGTTCCCTGGTGCA.........TCGGAGCCGATCCGGCTGCAG 6854

Scoring table: IDENTITY_NUC
            Gapop 10.0 , Gapext 1.0

Searched:   16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters:        32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
             Maximum Match 100%
             Listing first 45 summaries

Database :   EST:*
         1:  em_estba:*
         2:  em_esthum:*
         3:  em_estin:*
         4:  em_estmu:*
         5:  em_estov:*
         6:  em_estpl:*
         7:  em_estro:*
         8:  em_htc:*
         9:  gb_est1:*
        10:  gb_est2:*
        11:  gb_est3:*
        12:  gb_est4:*
        13:  gb_est5:*
        14:  em_estfun:*
        15:  em_eston:*
        16:  gb_gss:*
        17:  em_gss_hum:*
        18:  em_gss_inv:*
        19:  em_gss_pln:*
        20:  em_gss_vrt:*
        21:  em_gss_fun:*
        22:  em_gss_mam:*
        23:  em_gss_mus:*
        24:  em_gss_other:*
        25:  em_gss_pro:*
        26:  em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Score Length | DB | ID | Description | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 135.4 | 2.0 | 458 | 10 | AW501032 | AW501032 | UI-HF-BP0 | |
| 2 | 129.8 | 1.9 | 435 | 10 | AW501598 | AW501598 | UI-HF-BP0 | |
| 3 | 129.8 | 1.9 | 436 | 10 | AW501020 | AW501020 | UI-HF-BP0 | |
| 4 | 128 | 1.9 | 1798 | 17 | AG171124 | AG171124 | Pan trogl | |
| 5 | 125 | 1.8 | 430 | 10 | AW501038 | AW501038 | UI-HF-BP0 | |
| c 6 | 117.8 | 1.7 | 804 | 13 | BM617808 | BM617808 | 170006871 | |

```
c  7  117.2  1.7  AG171124  1798  17  AG171124  Pan trogl
c  8   111   1.6  BG809984  1516  12  BG809984  mgct002xd
c  9  109.2  1.6  AG030611  1341  17  AG030611  Pan trogl
  10  108.4  1.6  BG809984  1516  12  BG809984  mgct002xd
  11  108.2  1.6  AW064148   273  10  AW064148  SP0582 KR
  12  106.8  1.5  BG852363  1281  12  BG852363  1024034A0
  13  104.6  1.5  BH606535   828  17  BH606535  BOHAT59TR
c 14  104.6  1.5  BG809816  1651  12  BG809816  mgct001xk
c 15  104.2  1.5  AG032943  1542  17  AG032943  Pan trogl
  16   103   1.5  BG809816  1651  12  BG809816  mgct001xk
  17  101.4  1.5  AG030608  1189  17  AG030608  Pan trogl
c 18  101.4  1.5  BG852363  1281  12  BG852363  1024034A0
  19  100.6  1.5  AG060116  1278  17  AG060116  Pan trogl
c 20  100.4  1.5  AW200635   682  10  AW200635  da23h03.y
c 21  100.4  1.5  AY109601  1308  11  AY109601  Zea mays
c 22   100   1.5  AG077201  1309  17  AG077201  Pan trogl
  23   99.6  1.5  BM458211  1339  13  BM458211  AGENCOURT
  24    99   1.4  AG057392   997  17  AG057392  Pan trogl
  25    99   1.4  AK011555  1595  11  AK011555  Mus muscu
  26   98.2  1.4  AG162064   869  17  AG162064  Pan trogl
c 27   98.2  1.4  BM911414  1242  14  BM911414  AGENCOURT
  28    98   1.4  AV833062   633  10  AV833062  AV833062
  29    98   1.4  AQ893581  1205  17  AQ893581  HS_4832_A
  30    98   1.4  BG846745  1244  12  BG846745  1024014H0
  31   96.8  1.4  BQ922913  1169  14  BQ922913  AGENCOURT
c 32   96.6  1.4  AG030607  1538  17  AG030607  Pan trogl
c 33   96.2  1.4  BQ678719  1288  14  BQ678719  AGENCOURT
  34   95.8  1.4  BQ892090   946  14  BQ892090  AGENCOURT
  35   95.6  1.4  AG032943  1542  17  AG032943  Pan trogl
c 36   95.4  1.3  BJ037676   541  13  BJ037676  BJ037676
  37   94.8  1.4  AG032383  1169  17  AG032383  Pan trogl
  38   94.8  1.4  AG060256  2354  17  AG060256  Pan trogl
c 39    94   1.4  AL249930  1016  17  CNS01LU9  Tetraodon
c 40   93.8  1.4  AG060010  1201  17  AG060010  Pan trogl
c 41   93.6  1.4  BH512739   669  17  BH512739  BOHMX80TF
c 42   93.6  1.4  AG041123  1124  17  AG041123  Pan trogl
  43   93.4  1.4  AG039481  1299  17  AG039481  Pan trogl
c 44   93.2  1.4  AG043477  1625  17  AG043477  Pan trogl
  45    93   1.4  AG043469  1250  17  AG043469  Pan trogl
```

ALIGNMENTS

```
RESULT 1
AW501032
LOCUS        AW501032                458 bp    mRNA     linear   EST 01-MAR-2000
DEFINITION   UI-HF-BP0p-ais-f-07-0-UI.r1 NIH_MGC_51 Homo sapiens cDNA clone
             IMAGE:3072924 5', mRNA sequence.
ACCESSION    AW501032
VERSION      AW501032.1  GI:7114181
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 458)
AUTHORS      NIH-MGC http://mgc.ncl.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..458
```

## Left column

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3072924"
/clone_lib="NIH_MGC_51"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(4.4-7.4kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT      95 a      157 c      140 g      66 t
ORIGIN

Query Match      2.0%;    Score 135.4;  DB 10;   Length 458;
Best Local Similarity 60.4%;  Pred. No. 5.1e-11;
Matches  264;  Conservative   0;  Mismatches  161;  Indels  12;  Gaps  2;

```
Qy  3613  CGACAGGGTGCCTTCACCCGACCAACGTGCTGGGCACCCAGGTCCTGCTCGACGCCGC  3672
Db    25  CAACCCCGAGCCGTTCCTGCACACCATGTGGTGGGCACCTTCACCATTCTCGAGGCGGT  84

Qy  3673  GCTCCGCCACGGTGCGCCACCTTCGTCGCACGAGGTGTACGGCTCCCT  3732
Db    85  GCGACGCAACGGGGTGCCGGC---TGCATCACATCTCGACCGACGAGGTGTACGGCGACCT  141

Qy  3733  CCCGCACGGGGCCCCGGGAGAGCGAC-------CCCTGCTTCCGACCTCGCCGTA  3783
Db   142  CGAACTCGACGACCCCAGCCGTTCACCGACCACCCGTACACCCGTCGAGCCCATA  201

Qy  3784  CGGCGGTCGAAGGCGGCCTCGGACCTCATGCGCGTCGGTGCGCCACCACCCACGGCCT  3843
Db   202  CTCGGCGACCAAGGCCGGACGACATGCTGCTGGTGCGCCGCGTCTGGGTGCGGTCCTACGGCGT  261

Qy  3844  GGACGTCGGGGTGACCCGTGTTCGGCGACCTCGACGAACAACTTCGGCGCCCCACCATCCCGAGAAGCT  3903
Db   262  GGCGTGCGACGCTGTCGAACTGCTCCAACAACTACGGCCCGGTCCAGCACCATCGAGAAGTT  321

Qy  3904  CATACCGGCGCTTCCTGACCAGCCTCCGTCCCGTCGCCGCACCGTTCCCCTCTACGGCCGACGG  3963
Db   322  CATTCCGGCCCCAGATCACCAACGTGCTCACCGGACGGCGATGCAAGCTCTATGGCAGCGG  381

Qy  3964  GCGGCACGTGCGGCACGTCGGCGACTGGCGTCGACGTCAGGGCCGTCGAACTCGTCCG  4023
Db   382  GCACAACGTGCGCGACTGGATACACGTCGAGGACCACCAACAGCGCGGTC  441

Qy  4024  CGTGTCGGGCCGGCCGCCGG  4040
Db   442  GGAGAAGGGCGAGTCGG  458
```

RESULT 2
AW501598
LOCUS       AW501598       435 bp     mRNA     linear     EST 01-MAR-2000
DEFINITION  UI-HF-BP0p-ajg-d-07-0-UI.r1 NIH_MGC_51 Homo sapiens cDNA clone
            IMAGE:3074340 5', mRNA sequence.
ACCESSION   AW501598
VERSION     AW501598.1  GI:7115280
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 435)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

## Right column

cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES             Location/Qualifiers
source               1..435
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3074340"
                     /clone_lib="NIH_MGC_51"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /note="vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (4.4-7.4kb). Directionally cloned. Cells provided by
                     Louis M. Staudt, Ph.D. Library preparation by Maria de
                     Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT      90 a      153 c      129 g      63 t
ORIGIN

Query Match      1.9%;    Score 129.8;  DB 10;   Length 435;
Best Local Similarity 61.1%;  Pred. No. 3.6e-10;
Matches  250;  Conservative   0;  Mismatches  147;  Indels  12;  Gaps  2;

```
Qy  3613  CGACAGGGTGCCTTCACCCGACCAACGTGCTGGGCACCCAGGTCCTGCTCGACGCCGC  3672
Db    25  CAACCCCGAGCCGTTCCTGCACACCATGTGGTGGGCACCTTCACCATTCTCGAGGCGGT  84

Qy  3673  GCTCCGCCACGGTGCGCCACCTTCGTCCACGAGGTGTACGGCTCCCT  3732
Db    85  GCGACGCAACGGGGTGCCGGC---TGCATCACATCTCGACCGACGAGGTGTACGGCGACCT  141

Qy  3733  CCCGCACGGGGCCCCGGGAGAGCGAC-------CCCTGCTTCCGACCTCGCCGTA  3783
Db   142  CGAACTCGACGACCCCAGCCGTTCACCGACCACCCGTACACCCGTCGAGCCCATA  201

Qy  3784  CGGCGGTCGAAGGCGGCCTCGGACCTCATGCGCGTCGGTGCGCCACCACCCACGGCCT  3843
Db   202  CTCGGCGACCAAGGCCGGACGACATGCTGCTGGTGCGCCGCGTCTGGGTGCGGTCCTACGGCGT  261

Qy  3844  GGACGTCGGGGTGACCCGTGTTCGGCGACAACTTCGGCGCCCCACCATCCCGAGAAGCT  3903
Db   262  GGCGTGCGACGCTGTCGAACTGCTCCAACAACTACGGCCCGGTCCAGCACCATCGAGAAGTT  321

Qy  3904  CATACCGGCGCTTCCTGACCAGCCTCCGTCCCGTCGCCGCACCGTTCCCCTCTACGGCCGACGG  3963
Db   322  CATTCCGGCCCCAGATCACCAACGTGCTCACCGGACGGCGATGCAAGCTCTATGGCAGCGG  381

Qy  3964  GCGGCACGTGCGGCACGTCGGCGACTGGCGTCGACGTCAGGGCCGTC  4012
Db   382  GCACAACGTGCGCGACTGGATACACGTCGAGGACCACCAACAGCGCGGTC  430
```

RESULT 3
AW501020
LOCUS       AW501020       436 bp     mRNA     linear     EST 01-MAR-2000
DEFINITION  UI-HF-BP0p-ais-d-10-0-UI.r1 NIH_MGC_51 Homo sapiens cDNA clone
            IMAGE:3072834 5', mRNA sequence.
ACCESSION   AW501020
VERSION     AW501020.1  GI:7114157
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 436)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

```
ORGANISM   Pan troglodytes
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      BAC end sequences of Library RPCI-43
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 1798)
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
           and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
           1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
           Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
           end was generated during the R&D process and may have higher chance
           of clone tracking errors.
FEATURES             Location/Qualifiers
     source          1..1798
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      0 a    778 c    655 g      0 t    365 others
ORIGIN

Query Match      1.9%;      Score 128;  DB 17;    Length 1798;
Best Local Similarity  37.8%;  Pred. No. 7.4e-10;
Matches  614;  Conservative   0;  Mismatches  985;  Indels  27;  Gaps  6;

Qy  3911  CGCTTCCTGACCAGCCTCCGTGTCCGGCGGCACCGTTCCCCTCACGGCGACGGGGCGGCAC  3970
          | |  || | | |   |  || |  |     | |  ||  | |  |  |  |    |
Db    37  CCCNNCCGGGGNNNNCCCNNCCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCC    96

Qy  3971  GTGCGCGACTGGCTGCACGTCGACGACCACGTCAGGGCCGTCGAACTCGTCCGCGTGTCG  4030
          |   |  ||   |  |  | |  | |  | |  | | |   |    |  |  |   |
Db    97  NNNNNNNNNNNNNCNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN   156

Qy  4031  GGCCGGCCGGGGAGAGATCTACAACATCGGGGGCGCACCTCGCTCGCGGCGCCCGCGGCTC  4090
              | |   | | |  ||  |  |  | || ||  ||   |  |  |  | | |
Db   157  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNN   216

Qy  4091  ACGCACCCGTTGCTCGCACTGTCGCGGCCGGCCGGCGGCGGCGGCATCGTCCAGTCGAGAAC  4150
          | |  |   |  ||  | |   | | | | |  | | |  | | | |  |  |  |||
Db   217  GCGCGGCGGCGCNGCGGCCGGCGCGGCGGCGGCGGCGGCCCNNNNCCGCGGGG   276

Qy  4151  CGCAAGGGGCACGACCCGCGCTACGCGTCGACCACAGCAAGATCACCGCGAACTCGGT  4210
          ||            | |     |     |  | |   |    |  | | |  ||  ||
Db   277  CGGNNNCNCNNNNNNNNNNNNNNNNCNCNNNNNNNNNNNNNNNNNNNNNNCCGCGCC   336

Qy  4211  TACCGGCCGCGCACCGACTTCGCGACCGCGTGGCCGACACCCGAAGTGGTACGAGCGG  4270
               || |       | |   |||| |
Db   337  NCCGCNCCGGCCCCCCCCGCCCGCCCCGCCGCCGCCCCGCCGGCCGNNNCCGGCGCNC  396

Qy  4271  CACGAGGACGTGGTGGCGTCCCCTGCTCCGCGGCGGACATGACGTGGGCCGGACGGGG  4330
          | | |     | | || |  |   ||  |  | |      |      ||  |   |
Db   397  CCCGCGCGCCGCCNCGGGGCNCNNCCGCCCCCCCGCCCCGCCCCGCCGCCCCCCCCC  456

Qy  4331  CCGGCCCCGGCCGGCACACCGCGCGCCCCGGCCGGGCCGGTCGGCGTCGGCGTCCGTGAG  4390
          | | ||| | |  || |    |  |    | | |  |||  |  |  | |  |  ||| |
Db   457  CCGCCCCCGGNNGGNGGGNGGGGGGGNGGGGNNCGGNNNNCGGNNNNNGGGGGGGCGCG  516
```

---

```
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Eco RI site shown at the beginning of the sequence.
           Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
           cDNA Library Preparation: M.B. Soares Lab
           cDNA Library Arrayed by: M.B. Soares Lab
           DNA Sequencing by: M.B. Soares Lab
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
           Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..436
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3072834"
                     /clone_lib="NIH_MGC_51"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (4.4-7.4kb). Directionally cloned. Cells provided by
                     Louis M. Staudt, Ph.D. Library preparation by Maria de
                     Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D. "
BASE COUNT     90 a    153 c    130 g     63 t
ORIGIN

Query Match      1.9%;      Score 129.8;  DB 10;    Length 436;
Best Local Similarity  61.1%;  Pred. No. 3.6e-10;
Matches  250;  Conservative   0;  Mismatches  147;  Indels  12;  Gaps  2;

Qy  3613  CGACAGCGGTGCCTTCACCCGCACCAACGTGGCTGGGCACCCAGGTCCTCGACGCCGC  3672
          | | | ||   || |  ||  |  |  ||   ||  |  | |  || ||| |  ||
Db    25  CAACCCGAGCCGTTCCTGCACACCACCATGTGGTGGCACCTTCACCATTCTGAGGCGGT   84

Qy  3673  GCTCCGCCACGGGTGCGGCCACCTTCGTGCACGTCTCCACGCGAGGTGTACGGCTCCCT  3732
          ||    |  | |  || | | |  | |  |   | | | ||||||||||| | |
Db    85  GCGACGCAACGGGGTGCGGC---TGCATCACATCTCGACGACAGGAGGTGTACGGCGACCT   141

Qy  3733  CCCGCACGGGTGCGGCGGGAGAGCGAC--------CCCTGCTTCGACACCTCGCCGTA  3783
          ||| | |  |    | | | | |             |||| | |  ||| | |
Db   142  CGAACTCGACGACCCCCAGCGGTTCACCGGAATCCACACCCGTCGAGCCCCATA  201

Qy  3784  CGCGGCGTCGAAGGGGGCCTCGGACCTCATGGCGCTCGCCGCCACCACGGCCT  3843
          |  | | |     | | || | | |     |  ||   |  ||||| |  |
Db   202  CTCGCGACCAAGGCCGGCAGCGGCGACATGCTGGTGCGCGCCGCCTGGGTCGGTCCTACGGCGT  261

Qy  3844  GGACGTCCGGTGACCCGCTGTTCGAACAACTTCGGCCCCACCAGCATCCCGAGAAGCT  3903
          | | |  |   | |  || |   || | | |  |   |  |  | | |  |
Db   262  GGGTGCGACGCGTGTCGAACTGCTCCAACAACTACGGCGCGCCGTTCCAGCACATCGAGAAGTT  321

Qy  3904  CATACCGGCGGCTTCCTGACCAGCCTCCTCGTGTCGCGGCACCGGTTCCCCTTCACGGCGACGG  3963
          |  | | ||| |  | | |||  | | |    | | | | | | |  ||| |  |
Db   322  CATTCCGCGGCCAGATCACCAAACGTGCTCACCGGACGGCGGCGATGCGGATGCCAGCGG  381

Qy  3964  GCGGCACGTGCGCGACTGGCTGCACGTCGACGACCACGTCAGGGGCCGTC  4012
          ||             | ||| |  |  |  | | | || ||  || |
Db   382  GCAACACGTGCGCGACTGGATACACGTCGGAGGACCACCACAGCGCGGGTC  430
```

RESULT 4
AG171124
```
LOCUS       AG171124            1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-040F09.TJ.
```

```
Db   1580 GGCGCCCGNCGCNCGGCGGCNC-CGCCCCGCGCGCNCNCGGCGCGCGCNCGNNCGCGGCGCGCC 1638
               |||| || ||||| | | | |||||||  |||||||||| ||||||| || ||||||| ||
Qy   5522 GATGGC 5527
              | | |
Db   1639 GCCGCC 1644
```

RESULT 5
LOCUS       AW501038                      430 bp    mRNA    linear    EST 01-MAR-2000
DEFINITION  UI-HF-BPOp-ais-g-09-0-UI.r1 NIH_MGC_51 Homo sapiens cDNA clone
            IMAGE:3072976 5', mRNA sequence.
ACCESSION   AW501038
VERSION     AW501038.1    GI:7114192
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: M.B. Soares Lab
            cDNA Library Arrayed by: M.B. Soares Lab
            DNA Sequencing by: M.B. Soares Lab
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3072976"
                     /clone_lib="NIH_MGC_51"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85".
                     /lab_host="DH10B (LTI)"
                     /note="Vector: pTT3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (4.4-7.4kb). Directionally cloned. Cells provided by
                     Louis M. Staudt, Ph.D. Library preparation by Maria de
                     Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT    90 a    145 c    127 g    68 t
ORIGIN

Query Match             1.8%;    Score 125;  DB 10;  Length 430;
Best Local Similarity  60.4%;    Pred. No. 1.9e-09;
Matches  247;    Conservative    0;    Mismatches 150;    Indels 12;  Gaps  2;
```

```
Qy   3613 CGACAGGCGGTGCCTTCACCCGCACCAACGTGCTGGCACCCAGGTCCTGCTCGACGCGCCGC 3672
               |  || | | |||| ||||| |||||||| || ||||||||||  |||| || || |||
Db     25 CAACCCCGAGCCGTTCCTGCACACCAATGTGTGGCACCTTCACCATCTCGAGGGCGGT 84

Qy   3673 GCTCCGGCCACGGTGTGCGCACCTTCGTGACGCTCTCCACCGACGAGGTGACGGCTCCCT 3732
               |  ||| | ||||||||||| ||||| |||||||||||||||||||||||||||||| |||
Db     85 GCGACGGCAACGGGGTGCGGC---TGCATCACATCTCGACGACGAGGTGACGGCGCACCT 141

Qy   3733 CCCGCACGGGGCCGCCGCGGAGAGGGAC--------CCCCTGCTTCCGACCTGCCGCGTA 3783
               ||| |   | | ||| |  ||         ||    |||   || ||  ||||||
Db    142 CGAACTCGACGACCCCCAGCGGTTCACCGAATCACGCCGTACAACCCGTCGCGAGCCCATA 201

Qy   3784 CGCGGCGGTCGAAGGCGGCCTCGGACCTCATGGCGCCACCGCGGCCT 3843
               |||||| |  ||| |||  ||| | | |||||||| |||| ||||| |
Db    202 CTCGGGCGACCAAGGTCGGCAGTGACATGTGACTACGTCGTGTGCGGCGGTTGCGGGTGT 261
```

```
Qy   4391 CCGGGGCGGCCGGCGCCCCGGGCGGCGGTGGACCCCGGACCACCAGTTCCGGCA 4450
               | | | ||| ||||| |  | | |
Db    517 GGGGNGGCGGGGGGGNNNGNGNGGGGNNGGGGGGNGGGGGGNG 576

Qy   4451 TGAAGACGAATTCGGTGCCGCGCGGTTCCGCTCATCTTCCTCCAGCAGTGCTCCA 4510
               ||  | |  |||| | ||  | |||||
Db    577 GGGGCNGCGNGGGNGGGGGGGGGNNNGGGGCGGGNNGGGG 636

Qy   4511 CGGCGACCTGCCCCATCGCCTTGACGGGCTGTCTCATGGTGTGTCAGGGGAGGGGTCGGTGA 4570
               ||  |  ||||| | || ||| ||||||||| |||||| || |||||
Db    637 GGGNGGCGGGGCGCCCGCGCCGGNCGGNCCCGGGCGCGCCGGCGGCGCGGN 696

Qy   4571 AGGCCATGAGCGGCGAGTCGTCGAAGCGCGACCACCGGAACCGTGAGAC 4630
               || | | ||| || | ||  | |  |  ||
Db    697 CGNGGCGGCGNGGCGGGNGGGGCGCGGCGCGCGCGCGCGCGGGCGCGN 756

Qy   4631 CCCCGCGGCGGCGCCCCGCGCGCCGAGGGCCCATCATGTCGTGGCCACATGACGG 4690
               ||| | | | ||||||||| |  | | | |
Db    757 CGGCGNCGCGGCGCCGGNCGCCCGGNCCGNCGNCGGNCGGGNCGGCGCGC 816

Qy   4691 CGGTGCAGCCAGTCGATCAGCGCGACGCGGCG------GCCTGGCCCCCTTCCAGGGAG 4746
               || || |  ||| ||  ||||||||||||||
Db    817 NNNGCGCGGCGCGCGNCGGCGCGCGGNCGGGCGCGNCGCCNCGCCGGCNGCCGGCG 876

Qy   4747 AACAGCGAGTGCTCCACGAGCTCCTCGGACTCCCGCGCCGACACTCCCAGTTGCTCCCGC 4806
               ||| |||| |  | |  |  |||| ||| |
Db    877 CGGCGNCGGCGCGCGCGGCGGNCGGCGCGNNCGGGCCGCCGGCGCCNCNGG 936

Qy   4807 ACGCCGGCCCGGAACCCCTCGATCTTCCGCTGCACCGGCACGAAGCGGGCGGCCCGACG 4866
               || | |  ||| ||  ||  ||  | |||| |||
Db    937 CCGCNCGCGCNNNCGCGCCCCCGCGCGCGCNCNCGCGCCGGCGCGCGCGGGGGCGG 996

Qy   4867 GCGAGGCCGCAACGCGCTCGTGCCCACAGCTCGCGCCAGTGTCGCCACGGCCAGGCGCATCGCG 4926
               |||  | | | | ||||  |||
Db    997 CGGCGNCGGCCGCCCGCNCGGCGCGCGNCGCNGCCNNNNNNGCGCCGCGCGCGCGCG 1056

Qy   4927 GCCCGGTCGTCCGGGGAGACGAAGGG--TGCCTCGATCCGGGGGCGGAGAACCCGTTCACGA 4984
               ||| |  | |  | |  ||| |  |     ||| |
Db   1057 CGGCCGNGCGCGGCGCGCNNNCGCGNGGCCCCGCGCGCCCGGCGCGCGCGCGCCC 1116

Qy   4985 GGACGAAGGGCACCTGCCGCTCGTGCAGCGGCGGCGCGCCGTACCGTCGGGTTGGTGT 5044
               || | | | | |||||| |     | |  | | ||  ||  |  |||||| || || |
Db   1117 GNCCGGGGGNCCGGCGCCNCGCCCCG------CCGGCGGCNNCGGGGGCGCGCGGC 1170

Qy   5045 CCGGCGTGACGTCCGGAGACGAAGAGATGATGCCGGACCACCCCGGTCCACGAGCATCTCCG 5104
               || || ||  |  |  | |  | |  | | | || |
Db   1171 GCGGCGGCNCGGCGGCGGGNNCGCCCGCNNCGCGCCGCCGCGCCGCGCGCGCG 1230

Qy   5105 TGAGTTCGTCTCGGTCGAGC---CGCCCGGGGTCTGCGTGGCGAGCACGGGCGTGTAGC 5161
               |   | || || ||||  |      | | | ||||| |
Db   1231 CGCCCGCGCGCGGCGGGCGCNCGGCGCGCGNCGGCGCGCGCGCGCGCCGCGCGCGC 1290

Qy   5162 CCTGACGCGTGCGGCGCTGCCCCATCACCTGGCGGGGAAGAAGGGGTTGTCCA 5221
               || | |  |  |||  |||||  | |||  |||  |||| | |
Db   1291 GCCCCCGCGCGNCGCGCGGCCGGGCGGGGCCGCGCGGGGGCGCGGCTGCTCGTAGC 1350

Qy   5222 GTTCGGGGGTGACCAGTCCGACAGCTCGGCCGGCGCCGGGCGCCGGCGCCGCCGCCGCCGCGCTAGC 5281
               ||   | | | || || ||  |  ||||  ||| |
Db   1351 GCGCGGGGCGGCGGGCGCCGGCGCCGGCCGCGCGCGGGGCGCGGGGCGCGCG 1410

Qy   5282 CCAGCGCGTCCAGTGCGTCAGCACCGAGTCGCGGGTGCCGGTGGCCACACCGCCGCAC 5341
               |  |  | || |  ||||  |||
Db   1411 CCNCCGCGCGCGNCGCGGCG---------GNCGNCCGGCGCCCCCGGCCGCCGCCGCGCGC 1459

Qy   5342 CGTTCAGCACCCGGCTGACCGTGGCCTTGCTGCACGCCCCGGCTGCGATGTCGGCGA 5401
               ||  | |||| |    ||| |
Db   1460 GGGCCCGCGCGCGACGACCCCCAGCGGTTCACCGGCGGCCGCGCGNNNCNCNNCCGCGCGC 1519

Qy   5402 GCCCCATGGTCATGGCAACGCACTCTACCTGTCGGGGCGTCAGGGCGTGCCCACCGCGCG 5461
               ||  || | |||| |||  | | |
Db   1520 NCCCGCGCGCGCCGCNNCGCGCNCNCACCGGNGGNCGCGCGGCGCGGCGCCGCGCGGCC 1579

Qy   5462 CGGAACCGGCGGACTGCGGGGACGCGGGCGTCCGCGCCCCCCACGGGAC 5521
```

```
Qy   1042  TCGACGACCTCCAGCGTGCCGCGGCGGCGGGGCGGGATCCCCAGTGGTGGGTTGCGGTCACGTC 1101
            |||  || |||||||||||  ||| ||||  |  ||| ||||||||||| |||||||
Db    313  TCGATCACGCGCCATGACCCGCGTTCGCAATCGCGACACCGGATTGCCACCGTACGTG 254

Qy   1102  GAGGCGTACGCCCCGGGTGGCCGCCTCGCCGCGTCCCAGCTTCCGGCGCTCCGGCCAGC 1161
            | |||  ||   || |  |||| |  || | |  | | |||  |||  ||
Db    253  TTGAAGTAGCAGACGCCGTGGCCGGCCGCAAAGCTTTCCGCAATCTCGGGCGTAGTGACGACC 194

Qy   1162  ACGGCGAAGGGGAATCCGCTCGGCGGTGCCCTTGGACAGCATCGGCCAGGTCCGGCTCGATG 1221
            ||  |||||||||||| ||  |||||||| ||  ||  | ||||| |||  |||||||
Db    193  GCCCCGACCGGATGCGCGGTGCCCATCGGTTGCCCACCGTCACTATGTCGGCACGATG 134

Qy   1222  CCGAACAGTTCGCTGGCGGAGGAAGGCGCCGGCCGCCCCGCGCGGGTGAGGAGGACCTCGTCG 1281
            |||  |||| |||||||||||||||  | | |  ||||||  | | ||| |||||||  | ||
Db    133  CCGTGCGGCTCGAACGCCCAGTAGTGCCGTGCCGATCCCGCCGAAGCCGACCTGCACCTCG 74

Qy   1282  GCGACGAGCAGCACGCCGCGTCCCGGCAGGCGCCGGCGGATCCGCTCCGTCCAGTAGCCGGGG 1341
            ||   |||| |||||||| ||||||  ||| |  ||||| | ||  |  ||  |||  |||||
Db     73  TCCCGGATCGTTACGGCCGCGGCCGCCTTGCGTCGTACCGCGTGTACACCTTCTCGGACGCGTGG 14

Qy   1342  GGCGGCACG 1350
            | ||||| |
Db     13  GCGGACGCG 5
```

```
RESULT 7
AG171124/c
LOCUS       AG171124            1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-040F09.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
            PRIMERS
              Sequencing: TJ
            LIBRARY
              Vector    : pBACe3.6
              R.Site 1  : EcoRI
              R.Site 2  : EcoRI
FEATURES             Location/Qualifiers
     source          1..1798
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9558"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      0 a    778 c    655 g      0 t    365 others
ORIGIN
```

```
Qy   3844  GGACGTCCGGGGTGACCCGCTGTTCGAACAACTTCGGCCCCCCACCAGCATCCCGAGAAGCT 3903
            |   | || ||||  | ||  | | ||||||| ||| || |   ||||||||  |||
Db    262  GCGTGCGACGCTGTCGAACTGCTCCAACAACTACGGGCCGTTCCAGCACATCGAGAAGTT 321

Qy   3904  CATACCGCCGCTCCTGACCAGCCTCTGTCCGCGGCGACCGTCCCCTCCCTCACGGACGACGG 3963
            ||| |||||||   | ||| |  | ||  |  || | |  ||| ||| ||| |||  ||
Db    322  CATTCCGCCGCCAGATACCAACGTGCTCATCGGACCGTGATGCAAGCTCTATGGCAGCCGG 381

Qy   3964  GCGGCACGTGCGCGACTGGCTGCACGTCGACGACCACGTCAGGGCCCGTC 4012
            ||  ||||||||||||||| ||||||||| ||||| |  |||| |  |||
Db    382  GCACAACGTGCGGCGACTGGATACAACGTCGGAGGACCCAACAGCGCGGTC 430
```

```
RESULT 6
BM617808/c
LOCUS       BM617808             804 bp    mRNA    linear   EST 25-FEB-2002
DEFINITION  1700068716009 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
            1960044970929 5', mRNA sequence.
ACCESSION   BM617808
VERSION     BM617808.1  GI:18916030
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   Holt,R.A.,Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R. Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 240453151
            Fax: 240453580
            Email: HoltRA@celera.com
            Plate: NU01004AXP row: I  column: 23
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..804
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="1960044970929"
                     /clone_lib="A.Gam.ad.cDNA.blood1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                     hours after human blood feeding. cDNA inserts >500 bp
                     cloned directionally into pSport 1. Not 1 site is 3'.
                     Clones available through the Malaria Research and
                     Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT    151 a    251 c    284 g    118 t
ORIGIN

Query Match      1.7%;  Score 117.8;  DB 13;  Length 804;
Best Local Similarity   52.6%;  Pred. No. 2.4e-08;
Matches  257;  Conservative   0;  Mismatches 232;  Indels    0;  Gaps    0;
```

```
Qy    862  ACGGCCCGGCGGGCGTCTCGGGGGGCGGGTGCCCCGCGACGGCCGTCGGTGACGAACTCCAGG 921
            ||| ||||| ||| |  ||  ||||||  ||||||| || ||| || ||| ||  |  |||
Db    493  ACCGCCTTGGCCGGCGGCGGGGTCGCCGCGGTCGGCGTCGGGGAAGCCGGACTTCCAGGGCG 434

Qy    922  CCCCACAGCAGTCCGAGGCCCGCGTCACCTGGCCGAGCTGGGGCGGAAGCGGGACTGTCCAGGGCG 981
            || |  | | |||  | |||| | || | ||||||||||| || |  |  |||
Db    433  CCCACGAACAGGCCGGTGCCGCGCACATCGCCCACCAGCTGATACTCGTGGAGTATCGTGTACTGGATC 374

Qy    982  CGCAGCCGGCTCCTCGGATGAGCTCGGCCGCGGAGGACGGCCGCACGCGGTCGATCAGCCGGTCCGC 1041
            |||  ||||| | | ||| |||||  |||  |  | | ||||  |  ||   |  |||||  || ||
Db    373  CGCGACTGCTCCAGCAGGTAGCGGCCGCGCAGGTAGTGCCGAGCGCGGTTCTCCTGCAGCCGCTCCTCG 314
```

```
Query Match    1.7%;   Score 117.2;   DB 17;   Length 1798;
Best Local Similarity   41.5%;   Pred. No. 3.2e-08;
Matches 656;   Conservative   0;   Mismatches 900;   Indels 26;   Gaps 10;

Qy    379  GGGTACCGACCAGAGTCCCACCAGAGCTGGCTGCCGGGGGGCCGCGGGGGCGGGGCGGGGGGCGGCGGGGT   438
                || ||  |||   ||| |||||  | ||   ||||| ||  ||||| || || |||||
Db   1797  GCGNGGGCCGGGGCCGGGCGGGGGGGGGCGCGCGGGGCGGCGGGGGCGGCGGCGGCGGCGGCGGC  1738

Qy    439  GGGCTTGCCTGCCGGTGCCGGTAGGGCTCAGCCCGGGACCCGCGACCGTCAGCCGGCCGGAGGT   498
             ||| ||  |||| || |  ||| ||  | ||  ||| ||| | |||| || ||
Db   1737  GCCCCGGGGGNNGGGGCGGCGCNNGCCGGGCNGCGGGGCGGGGGCGGGGGGCGGGGGGC  1678

Qy    499  GGGGTCATGATGCCGGTCAGCATCTTGATCGTG--GTCGACTTGCCGGCTCCGTTGGCG   556
              |||| |||||  | |||     ||  || |  ||| || ||| | ||||  |||
Db   1677  CGCGGCCGNGCGGCCGGCCGCCNGCCCGGTATCTCGAAGGAGCGTCGTCGACGGGCGGCG  1618

Qy    557  CCGATGTAGGCGGTCTTCGTGCCGCCGTATCTGAAGGAGGACGTCGTCGACGGGCGGC   616
             ||  ||  | || ||  ||  ||| || || ||||| || || ||
Db   1617  CGNGNCGCGGGGGCGGGNGCGGNCGGCGGGCGCGGGCCGCCGGGGGCGGCCGGGCCGCGC  1558

Qy    617  ACGACGCGGTACCGGCGGTCAGGAGGGCTGGAGAGGGCGCGGNNGCGGCGGGGGCCTCGCGT   676
              ||| | |||||| || |||  || |  | ||  |||| ||||
Db   1557  GCGGGNGNGGGCGGGCCGGGNCGGNCGCGGGGNGNGGNCGGGGGNGGGCCGGGGCCGCGGN  1498

Qy    677  TCGGCCAGCCGGAACTCCTTGACGAGGTGTTCGGCCACGATCACGCGGATCACCCGCTCGA   736
            ||| || || || ||| |   || ||| |||||| ||||  | ||  |
Db   1497  GNGGNCGGCCGGGCGCGCGGCCCGCGCGCGGNGC---CGNGNGGCGGC  1441

Qy    737  CGGGCCGTCTCCAGCAGGCGGCCAGGCGCCCTCGTCGAGGGTGAACGGCG   796
             ||| || | ||| ||  | |||| ||||| |  ||  | ||
Db   1440  CGGNCGNCCGCGCGGCGCCGCGNCCGGCNGCGCCGCGCCGCGCCGGCGCCCCGCG  1381

Qy    797  GTGCCAGCCGGATCGTGCCGCGCGCGCCCCGCCAGGGAGGTCGGC--AGCCCCAGTGCGCGCCGCCCGGCCG   854
            |||| || || ||  |  || | | || ||||  ||| |||||| || | ||| |||| ||
Db   1380  GCGCGCCGCGGCCGCCGGGGCGGTCTCCGGGGCGGTCGCCGCCGGGCCGCGCCGGCGCCCCGCGGCGCCG  1321

Qy    855  GGTGTAGACGGGCCCGGGGGTCTCGGGGGCGGTGTCCCCCGGGCCGGCGGTGCGTCGGTGACGAA   914
              || |   || ||| ||| |||  ||| |  | | || |||| ||
Db   1320  CGCGCCGCCGCGGCGNGCGGGCGNCGGGGCGGCGGGGCGGGGNCGGGGCGGGGCGGCGCGCGCNCNC  1261

Qy    915  CTTCCAGCCCCCACAGCAGTCCGAGGCCGGCTACCTGGCCCGAGCTGGGGGAAGCCGGGACTC   974
             ||||| |   ||| || || | || || |||| || |  ||||   |||| || |
Db   1260  CGGGCGCGGCNGNCGGCGNCGCGCGGCGGNCGGCGGGCGGCCGCGGGGCGGNNGCC  1201

Qy    975  CAGGGGCGCCAGCCGCTCCTCGATGAGCTCCGCCGAGGACGCGCACGCGGTTCGATCAGCCG   1034
             ||| |||| |||| | ||  ||  |  ||| ||| || |  || |
Db   1200  GCGGNGCCGCAGGCGCNCGCCGCCGCCGCNCGCGGC-CGCCGGCGCGGCGGNNGCGCG  1142

Qy   1035  GTCGCCGCTCGACGACCTCCAGCGTGGCGGCGGCGATCCCCAGTGGTTGCTCGC   1094
             ||| | || ||| ||| |  ||| |||  | |||  ||| ||  || ||| | |
Db   1141  CGGGNGGGCGNGGGCGGCGCCGGGCGGCGCGCGGGCGGCGGCGGNGNCGGGGC  1082

Qy   1095  GTACGTCGAGGCGGTACCGCCCCCGGGTGGCCCGCCCTCCGGCGCTTCCGCGCGTCC   1154
             |||  | ||  ||| | ||   |||  ||| | ||| |  |  ||
Db   1081  GCGNNNGCGCGGCGNCGCGCGGCGCNCGCGGCGCGGGCCGGC-GCNNNNGGCGCGGCGGCGGNGC  1023

Qy   1155  GGGCCAGCACGGCGAAGG-GGAATCCGCTCGCGTGCCCTTGGACAGCATCGCCCAGGTCCG   1213
             ||| | |  |||| ||| ||   |  || | | |  |||||   || |
Db   1022  NGCNGGGCCGCNGNGGGCGGGCGGCGCCGCGCGCGCGCNNNGCCCCGCCGGNGNGGGC  963

Qy   1214  GCTCGATCCGACAGTTCGCTGGCGAGGAAGGCGCCCCCCGGTGCCCGCCGCGGTCGAGGA   1273
             ||| || ||| || | |  ||| | ||| | || ||||  || || ||||
Db    962  GCGGGGCCGGCGGCCNNNGCGCC-CGNGCGCGGGCGCCGCGGCGGC---CGCGNGNGGCGGGGG  904

Qy   1274  CCTCGTCGGCGACGAGCAGCACGCGCCGTCGCGCCGGCGATCCGCTCCCAGT   1333
             || ||| ||| ||| ||  | |||||  ||  | |  || ||  |
Db    903  CGCGNGGCGCGGCGGCGGNGCGGCGGNGCGGCGGGNGNCGGTTCGAAGACCAGGGCCG  844

Qy   1334  AGCCGGGGGCGGCACGATGACCGCCTGCCGCCGGCGGNNNGCGGGNGNGGCGGCCGG   1393
             ||| | ||||| | | |||  | |  ||||| || | ||   || || |
Db    843  NCGCGCCGGCGNCGCGCGCCCGGCGGC---CGCGNGNGGCGGGGCG  788

Qy   1394  AGACGTTGGGCTTCTCCGACGATGCCGCGATCGCCGGCGGCCACGAGGTCGCGCAGCGTCGCCACG   1453
```

```
Db    787  CCGGCGGGGGCGGNCGGCGGNCGGCCCCAGGGGACACGGGCCCAGGGGGCAGGGGGGACACGGGCCGCCGCCGCGGCGCG   728
             |||| ||| ||||| | |  ||  | || | ||  | | ||| ||||   |||| ||   | ||  |||
Qy   1454  AGGGGTACTCCAGGCGCGCGGGGCGCGGGCCGGGGNGCGGGCGNCGGTTAGCCCGNCGGCCTGTAGCCAGCACGCTGT  1513

Db    727  CGGCGGGGGCCGGGGGGGGCGGGGGGGCGGGGGGGCGGCGGGGCCGCGGGGCCGGGCCCCCATGGTCT   668
             ||| || ||   |||  |||||   || |  ||| ||||| || ||||  ||   ||
Qy   1514  TGCCGCTGAAGGCCTGGTGCCGATGTCCCAGTGACCAGCATCCGGGCCGCGCGCCCCCGCCCATGGTCT  1573

Db    667  CGCCCCGCCGCGGCGGGGCCCGGCGGGGGCGGCAGATCCGGTTGCGCGCGGTTCGCCCCGCGCCGC   608
             || ||||||||  ||  || || | |||| |  ||  ||| | ||||
Qy   1574  TGCCGTGGAAGCGCGTGGCGCCGCAGGGGCGCAGATCCGGTTGCGCGCGGTTCGCCGGTCGCCCT  1633

Db    607  CNCCCNNNNCCCCCGCGCGCCGCNCGGGGNNNCCCCGNCC   548
             || ||||     |||  | || | |||  || |||| |
Qy   1634  GGACGACCCGCAGGGCGGCTTCGAACACCTCCGGACGGCGGTGGAGAAGAAGGCGTAGGTGT  1693

Db    547  CNCCNNNNCCNCNCNNNNCCCCCGCGCGCGCCGCGCCNNNNNCCGNNNNNCGCCNC   488
             || ||||    | || | ||||| ||  || |||||    |||| ||
Qy   1694  CGAGCTGTTCGGGCAGCAGCCTGGCAGCAGTTCAGCAGCAGCCGGCGGTCCGCGCTGG  1753

Db    487  CCCCCNCCCNNNNCCCNNNCCGGG--------GCGGGGGGCCGGGGCGGCGGCG   438
             ||||| |||   |||  ||
Qy   1754  CGCTGTCTGGACGTTCCACAGCGGCGGGGCCTGGGTGGTGCTGAGTGCCCTCGACGACCTCCG  1813

Db    437  NGC-GCGGGGCGGGGGGCGGNNGGGGCGCGGGCGGNGCCGGGGCGNGNGGCGG   379
              || || | || |||| ||  || |  ||| || | | |||
Qy   1814  GGTGCCGTGGCCCAGTGACTGGTGAGGGTCCGCGCCGCGAAGTCGAGGTACTGGTTGC  1873

Db    378  GCGGCCGCGCGCCGCGCCCCGGGGCGGCGGGGCGGGGCGGGGCGGCGCGTCCGTGGACGGG   319
             || ||| ||  |||  | | ||  | ||| | || |||
Qy   1874  CGTCCAGTCGGTCAGAACGGGACCGCGCTCGGCGAAGACCCGCGTCCCTCGGCGAAGACCGGT  1933

Db    318  NNNNNNNGGNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCCCGGNNNNNGGGCCCG   259
             ||||  ||||||||| |||||||   | |||||    || |
Qy   1934  CTTCCTCGGAGGCGCCCGCCGC  1955

Db    258  CCGCGGCGCCGCCGCCGCCGC   237
             ||  |  || || |||  |
```

RESULT 8
BG809984/c

```
LOCUS       BG809984    1516 bp    mRNA    linear    EST    22-MAY-2001
DEFINITION  mgct002xd11f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct002xd11f 5', mRNA sequence.
ACCESSION   BG809984
VERSION     BG809984.1  GI:14180964
KEYWORDS    EST.
SOURCE      Magnaporthe grisea
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES             Location/Qualifiers
     source          1..1516
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct002xd11f"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /dev_stage="Germinated conidia on appressorium-inductive
```

surface"
/note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
Site_2: XhoI; The appresorium formation-specific cDNA
library was constructed from conidia germinated for 5-8 hr
on an inductive surface. The library contains over 55,000
clones with average insert size of 1.5 kbp."

BASE COUNT    65 a    657 c    660 g    16 t    118 others
ORIGIN

Query Match    1.6%;    Score 111;   DB 12;   Length 1516;
Best Local Similarity   44.8%;   Pred. No. 2.7e-07;
Matches 538;   Conservative   0;   Mismatches 638;   Indels   25;   Gaps   6;

```
Qy  4102  GCTCGCACTGTGGGGCGGGGCCCGCGAGCGCATCGTCCACGTCCGAGAACCGCAAGGGGCA  4161
                ||||     ||---||||||| -|-|||-|-|||    || |-|-|||
Db  1418  GCGNGGGGCGGCGGCGGCGGGGCGGCGGGGCGGCGCGGCGTCGCGCNCGCGCGGG  1359

Qy  4162  CGACCGGCGGCTACCGCGTCGACCACAGCAAGATCACCGCGGAACTCGGTTACCGCCGCG  4221
                ||||||-|-||--|| |  |-|  ||    |||| |-|-|||||
Db  1358  CGCTGTGCGCGCTGCGCGCGGCGCGGCGGCGTGCGGCGGCGTGCGGNGNGNGNNCGCGGCGCG  1299

Qy  4222  CACCGACTTCGCGACCGCGGCTGCCGACACCGGCGAAGTGGTACGAGCGGCACGA---GGA  4278
                |||--|||--|| |||  ||--| ||-|||||||-|-|-||
Db  1298  CCCGGCGCGGCGGGCGGCGGCCGGGCGCGGTGCGCGCGGCGGGCGGCGCGGCGNCTCCCG  1239

Qy  4279  CTGTGGCTCCCCTGCTCGCCGCACATGACGTCGGGCGCCGGACCGGCCCAACCACCGCCCC  4338
                ||||--|--|| |  | ||
Db  1238  CCGCTGCNGTGCGCGGCGGTTCTGCGCGCGGGGCGTCCGGGCGGCGTCCGTGAGCCGGCGGC  1179

Qy  4339  GGGGCGCACACCGCGCGGCGGCGGCGGCGTGGGCGGGGGCGTCCGGTGGCGGGGCGCGGCGGC  4398
                |-|-|  ||  |-||||| |-|-|||  |||| |-|-|||
Db  1178  TGCGGCGGCGGCGGNGNGNNGNNCGCGCGGGCNCGGCGGNCNCNCGNGGGCGCGGCGGGCG  1119

Qy  4399  CGGGCGCCCCGGCGGGGCCGGCGGCGGTGGACCCCCGGACCACCAGTTCCGGCATGAAGACG  4458
                |||--|||||| |||-|  ||   |  ||| ||
Db  1118  CGGGCGGGCNCCGCCCGGGGCGGCGGCCNNCNCGGCGCGCCGGGCCGCGGCGGCGGCGGG  1059

Qy  4459  AATTCGGTGCGCGGCGGCGTTCCGCTCATCTCCTCCAGCAGTGCGTCCACGGCGACC  4518
                ||||  ||-|-|| | ||--|  |  |||  ||| ||-|-|-||
Db  1058  CGCGCGCGCGGCGCCGGCGTCCGGNCGGCTTCGCCGCGGCGGCGGGGCGGCGCCG  999

Qy  4519  TGCCCCATCGCCTTGACGGGCTG-TCTGATGGTGGTCAGGGGAGGGTCGGTGAAGGCCAT  4577
                || -|||| ||-||||||||  | |  | |||||||||  |-|-|| ||
Db  998   CGGGCGGCGGCGCGCGCGGCGCCCCGCCGGCGGGGGGGCGCGCGGGGGGCCGCGGGGGGCC  939

Qy  4578  GAGCGGCGAGTCGTCGAAGCGACCCGAGATGTCACCGGACCGTGAGACCCCCGCCG  4637
                || -|-||-|  ||||  |||| || || |-|-|-||
Db  938   CGGGCGGGCGGCGGCGCNCGGCGGCNCGGNGCCCGCGNGCGCGGGCGCCGCGCGGCGCCG  879

Qy  4638  GCGCGCCGCCGCACGGGCCGAGGGCNCGCGCNCGCGCCGCCATCATCGCTGGCGCACATGACGGCGGTGCA  4697
                |-|-|||||||| -|| |  ||-|-||||||  ||| ||||
Db  878   CNGCCCGCGGCGGGGGCCGCGNCGCCGCCGGNCCGCGGCCGCCTTGCGCGCCGCCNNCGC  819

Qy  4698  GCCCAGGTCGATCAGCGGCGACGGCGGCGGCGACGCGGCCCCTCCAGGGAACAGCGAGTG  4757
                ||| ||||||  |||| |  || ||||| |||||||
Db  818   GGCGCGCGCGCCGCCGGCGCCGGCGGCGNCGCCGGCGCGGCCNCCGCGGGNCGCGGCGGCC  759

Qy  4758  CTGCACGAGCTCCTCGGACTCCCCCGC-CGACACTCCCAGGTGCTCCCGCACGCGGGCCCC  4816
                ||| |||| |||||| |  ||||| |   | ||||  ||-|-||
Db  758   GCGCGGGGGGCCCCNNCCGGCNCCGCGGCNCCGGNGCCGGCGCGGCGCGCGGCCGGGCCGC  699

Qy  4817  GGAACCCCTTCGGATCTTCCGCTCACCGGCACGCAGGAGGCGGCGGGCCCGACGGCGAGGCCGA  4876
                ||| |||| |||  ||-| | -|  ||||||| ||-|-|||
Db  698   GCCGGCGGNCGNCNCCGGCNCGCCCGGCGGGGCGGCCCGGGCCGCCGGGCCGCGNGGCGGGGC  639

Qy  4877  CGCGGTCGTGCCCCAGGTCTCCCAGGTGCGGCCCACGCGGCCAGGCGCCAGCGGCCCCCGTCGT  4936
                |-|||||  ||   |   |  |-|-|||
Db  638   GGGCGCCGGGCCGGCGCGGCGCGCCGCGCGGCCCCCCGCGCCCCCNCCCCGGNCCCCGC  579

Qy  4937  CCGGGAGACGAAGGGTGCCTCGATCCGGGCGGAGAAACCCGTTCACGAGGACGAAGGGCA  4996
                ||| |  | || |-|-|   |  |  -|-|||
Db  578   CCCCNGCGCGCGCGCGCGCGCGGCGNGGCGGCGCGCGCGNCGCNGNGGNGNCCG  519
```

```
Qy  4997  CCTGCCGCTCGTGCCAGCCGGCCGTACCGTCCGGTCTCCGGCTGGTGTCCGGCCGGTGAGTC  5056
                ||--||--||  | || -|-|||
Db  518   CCCCCCCCCCGCGGGGCGCCGCNNCNCNCCGGCGCCGGCGGGGGCGGCGGCGGCGGCGGC----  463

Qy  5057  CGGAGACGAAGATGATGCGGACACCCGCGTCCACGAGCATCTCCGTGAGTTCGTCCT  5116
                       |||| -|-|||  ||| || |||||
Db  462   --------GCGCGGCGGCGNCGGCGGCGCGCGGGGNCGGCGGCGGCGGCGCGGCCGGCGCCGCNGCCG  413

Qy  5117  CGGTCGAGCCGCGGGTCTGCGTGGCGAGCACGGGCGTGTAGCCCTGACGGGCTGAGCG  5176
                |-|-|||-|| ||||| |-|-|| |  |||||
Db  412   CGGCGGCCCGCCCGNGGGNGCGNCGGGGNGNCCGGGCGGCCCGC---GCCCCGCGCCG  356

Qy  5177  CCTGCCCATCACCTGGGCCAGTGCGGCGGGGAAGAAGGGGTTGTCCAGTTCGGGGGTGACCA  5236
                ||| |||| |||  ||  |  |||| |-|-|||
Db  355   CGGGCCCGGCGCGCGGCGCCGCCGGCCGGGGCGGGCCCCG---CGGGCGCCCG  299

Qy  5237  GTCCGACCAGCTCGGCGGCTCGGCGCGTCGTCTGCGGGCTGCTCGGCGTCGTCGCAGCGGTCCAGTG  5296
                |-||| |  || ||||  |||||  ||| |  ||-|-|||||
Db  298   GCCCGGCGCCGGGCGGGGGGCGGGCGGGGNCGNCGGGCGGGNGCGGCGGCGCGCGCGCG  239

Qy  5297  C 5297
Db  238   C 238
```

RESULT 9
AG030611/c

LOCUS     AG030611     1341 bp     DNA     linear    GSS 01-NOV-2001
DEFINITION   Pan troglodytes DNA, clone: PTB-003A20.F, genomic survey sequence.
ACCESSION    AG030611
VERSION     AG030611.1     GI:16557484
KEYWORDS     GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
           BAC Library clone:PTB-003A20.F.
  ORGANISM    Pan troglodytes
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE    1
  AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE       BAC end sequences of Library PTB
  JOURNAL     Unpublished
REFERENCE    2  (bases 1 to 1341)
  AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE       Direct Submission
  JOURNAL     Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
           and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
           1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
           Tel:81-45-503-911, Fax:81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library PTB This BAC end
           was generated during the R&D process and may have higher chance of
           clone tracking errors.
           PRIMERS
             Sequencing: -21M13
           LIBRARY
             Vector    : pKS145
             R.Site 1  : SacI
             R.Site 2  : SacI.
FEATURES      Location/Qualifiers
     source       1..1341
             /organism="Pan troglodytes"
             /db_xref="taxon:9598"
             /clone="PTB-003A20.F"
             /sex="male"
             /cell_type="lymphoblast"
             /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      48 a     476 c     700 g     14 t     103 others
ORIGIN

Query Match    1.6%;    Score 109.2;   DB 17;   Length 1341;
Best Local Similarity   45.1%;   Pred. No. 5e-07;

Qy  4847  CGAAGCGGCGGGGCCCCGACGGCGGAGGCGAGCCGCGCTCGTGCCCCAGCTCCGGCCAGG  4902
                |  | ||||            ||||                    |||||       ||||
Db   229  CGGCCGCCGCCGCGGNNGGCGTGNNGGCCGNNGGCCCGCGCNCCGCGGCCGCNCCGCCCGG  174

RESULT 10
BG809984
LOCUS       BG809984                1516 bp    mRNA      linear   EST   22-MAY-2001
DEFINITION  mgct002xd11f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct002xd11f 5', mRNA sequence.
ACCESSION   BG809984
VERSION     BG809984.1  GI:14180964
KEYWORDS    EST.
SOURCE      Magnaporthe grisea
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG)
FEATURES             Location/Qualifiers
     source          1..1516
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct002xd11f"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /dev_stage="Germinated conidia on appressorium-inductive
                     surface"
                     /note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed from conidia germinated for 5-8 hr
                     on an inductive surface. The library contains over 55,000
                     clones with average insert size of 1.5 kbp."
BASE COUNT      65 a       657 c      660 g     16 t     118 others
ORIGIN

Query Match             1.6%;    Score 108.4;  DB 12;  Length 1516;
Best Local Similarity  43.0%;    Pred. No. 6.6e-07;
Matches 532;  Conservative 0;  Mismatches 698;  Indels 8;  Gaps 4;

Qy  4105  CGCACTGTGCGGCGGCCGGCGCCGCCGGCGCATCGTCCACGTCGAGAACCGCAAGGGGCACGA  4164
                |||| ||| ||||     ||| ||  ||||
Db   239  CGCCGGCGCGCCGGCCGGCGCCGCGGCGCCCCGCGCCCGCGGCGCCCGCGCGGGCGGGGGC  298

Qy  4165  CCGGGCCGCTACGGCGGTCGACCACAGCAAGAATACACCGGGAACTCGGTTACGGCGCCCAC  4224
            |  ||||          |||||||     |      |||||||
Db   299  GGGGCCGGCNGCGCCCGCGGCGGCGCGGCGGCGTACGAGGCGGCACGAGGACTGGTG  358

Qy  4225  CGACTTCGCGACCGGCGCTGGCCGACACCGGCGAAGTGGTACGAGCGGCGCGCCGGCNG  4284
             ||||| |          |||                    | ||||
Db   359  GCGCGGGGGCGGGCCGGGGCGCGCCCCGGGCGGNNGCCNGGCCCCCGACCCCGAGCCGCGG  418

Qy  4285  GGCGTGCCCTGCTCGCCGCGACATGACGTCGGGCGGCCAACCACCGGGCCCCGGGCCGG  4344
            || |                           |||
Db   419  CGGG---CCGGCGCCCCCCGCCCCCGCGGACTCCTGCGGCGCTCCGCGGACGTCCGCGGC  475

Qy  4345  CACACCGGCCCCGGGCGCGTGGCCGGGCGTCCGTGAGCCGGGCCGG  4404
             ||||| |||           ||||                    | |||
Db   476  CCCGGGCGGCGGGGGGGGGGNGNGNNGGGGNGGGGGCGGGCGGNCNCCNGCGCCCC  535

Qy  4405  CCCCGCGGGGGCCGGCGGGCCGGCGGTGGACCCCGGACCACCAGTTCCGGCATGAGACGAATTCG  4464

ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 273)
AUTHORS    Goh,S.-H., Park,J.-H., Lee,Y.J., Lee,H.G., Yoo,H.-S., Lee,I.-C.,
           Park,J.-H., Kim,Y.-S. and Lee,C.-C.
TITLE      Gene expression profile and identification of differentially
           expressed transcripts during human intrathymic T-cell development
           by cDNA sequencing analysis
JOURNAL    Genomics 70 (1), 1-18 (2000)
MEDLINE    20541704
COMMENT    Contact: Sung-Ho Goh
           Genome Center
           Korea Research Institute of Bioscience and Biotechnology
           Oun-dong 52, Yu Sung-Gu, Daejon 305-333, Republic of Korea
           Tel: 82-42-860-4473
           Fax: 82-42-860-4479
           Email: gohsh@mail.kribb.re.kr
           Seq primer: T7
           High quality sequence stop: 273
           POLYA=No.

FEATURES           Location/Qualifiers
     source        1..273
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_lib="KRIBB Human CD4 intrathymic T-cell cDNA
                   library"
                   /tissue_type="Thymus"
                   /cell_type="intrathymic T-cell"
                   /dev_stage="CD3+4+8- single positive stage"
                   /note="vector: pGEM-T; cDNA was made from total
                   cytoplasmic RNA of sorted human intrathymic CD3+4+8-
                   T-cell, adaptor ligated, amplified with PCR, and cloned
                   into pGEM-T vector."

BASE COUNT      55 a      92 c      82 g      44 t
ORIGIN

Query Match      1.6%;     Score 108.2;   DB 10;    Length 273;
Best Local Similarity   62.3%;   Pred. No. 6.2e-07;
Matches  170;   Conservative   0;   Mismatches 103;   Indels   0;   Gaps   0;

Qy   3809 CTCATGGGCGCTCCGCCACCACCGCACCCACGGCCTCGGACGTCCGGTGACCCGCTGTTCG 3868
              | ||| ||| |||| ||| ||| ||||||||| ||||||||||| |||| |||||||
Db      1 CACCTGGTCGCCGGCCTGGCAGCTGACCTATGGGCTCGGTGCTGTGATCACCAACTGCTCC 60

Qy   3869 AACAACTTCGGCGCCGCCCCACCAGATCCCGAGAAGCTCATACCGGCGCTTCCTGACCAGCCTC 3928
              | ||||||| || |||| | ||| | ||| ||| |||||| | ||||||
Db     61 AACAACTACGGCCCGTTCCACTTCCCGAAAAGCTGATTCCGCGTTGGTAATCCTCAACGGC 120

Qy   3929 CTGTCCGGCACCGTTCCCCTTCCCCTGTACGGCGACGGGCGCACGTGGCCGGCTGCAC 3988
              | || |||| ||| |||||||||||| || ||||| || || || ||||||||||
Db    121 CTGGCAGGCAAGCCGCTGCCGGTGTACGGCAACGGCCTGCAGGTGCGCAGGCTGGCTGGCTGTAT 180

Qy   3989 GTCGACGACCACGTCAGGGCGTCGAACTCGTCCGCGTGTCGGGGGGAGATC 4048
              ||| | | |||||| ||||| | ||||||||||||| ||
Db    181 GTCGAGGATCACCGCCGGGCCCTGTTCAAGGTGGTGACCGAAGGCCGCGTCGGCGGAGACC 240

Qy   4049 TACAACATCGGGGGCGGCCACCTGCTGCCCAAC 4081
              ||| || ||| || | ||||||| ||
Db    241 TACAACATCGGCGCCGCCCACAACGAGCAGAAGAAC 273

RESULT 12
LOCUS       BG852363
DEFINITION  1024034A03.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG852363
VERSION     BG852363.1  GI:14233547
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii
ORGANISM    Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

---

Db    536 CGGGCGGCCGCCGCCGGCGNCGCCGGGGCGCCGGCCGCCGNGCGGCGGGGGCGGCGCCGGG 595
              | |||| | ||| ||    ||| | ||||
Qy   4465 GTGCCGGCGGGCGTTCCGCTCATCTCCTTCCAGCAGTGCGTCCACGGCGACCTGCCCC 4524

Db    596 CGCCGCGGGGGGCGCCGGCCGGCGGCCGGNCGCCGGCCGCGGCCGCC--NCGGCCCC 653
              || | |||| || | |||| |  |||| |||
Qy   4525 ATCGGCCTTGACGGGCTGTCTCGATGGTGGTCAGGGAGGAGGGTCGTGAAGGCCATGAGCGGC 4584

Db    654 GGGCCGCGCGCCGCGGCCGGCGNGNCGCGGGGGGCCGGGGGGCGGCGGGG 713
              |||||| | |||| |||| | ||||| ||| |
Qy   4585 GAGTCGTCGAAGCCGACCACCGAGATGTCACCGGGAACGTGAGACCCTGCGGGGCCCGCG 4644

Db    714 CGGGGCGGCGGGGCGGCCGCGGCCGCNCGGCGGNNGGGGGCCCGNCGGCGNGCGGNGGCCGG 773
              ||| ||||| |  ||||| |||| ||| | ||| ||||||||| |
Qy   4645 GCCCGCCACGGCGGCGGCCGAGGGCGCCATCATGTCGCTGGGCCACATGACGGCGGTGCAGCCCAGG 4704

Db    774 GCGGCGGCGCNGGCCGCCGCGGCGNGGCGGCCGGCGCCGCGGCCGCCGCCGCCGCCCAGG 833
              |||||| || | ||| |||| | ||| ||||| ||| |||||
Qy   4705 TCGATCAGCGCGGACGCGGCGGCGGCCTGGCCCCCCTCCAGGGAGAAACAGCGAGTGCTCACG 4764

Db    834 CGGGGGCGGGGGGCGGGGGCGNCGGCCGGNCGGGGCGGCGGGGGGGCGCCGGGGGGCGCCG 893
              | ||| ||||  | | |||||| ||| | | |||| || ||| | | ||||
Qy   4765 AGCTCCTCGGACTCCCGCGCCCAGGTGCTC--CCGCACGCCGGCCCGGAACC 4822

Db    894 CGCGGCGGGCGGGGCGGCGNCGNCGCGGCGNGCCCGCGCCCCCGCGCCCCC 953
              | ||| ||||  |||| | ||| |||| | ||| |||||| |||||
Qy   4823 CCTCGATCTTCCGGCTGCACCGGGCGCAGGAAGCGGGGCGGGCCCGACGGCCGACG--CGC 4881

Db    954 GCGGCCCCCCCCCGCGGGGCCGGCGGGCGGCCGCGCGCCCCGGNC 1013
              | |||| | | ||||||||| | || | |||||||||||||||| |
Qy   4882 TCGTCCCCAGCTCCCGACCTCCCGCAGGTGCCGCCACAGCGCCCCCGTCGTCCGGG 4941

Db   1014 GCGGCCGGCGAGCCGNCCGACGCCCGGGCGCCGCCGCCGCCCGGGGC 1073
              | ||  |||| || | | ||| || || | || || |||| |||||||||
Qy   4942 GCGCCCCCGGCGAGCCGNCCGACGCCCGGGCGCCGCCGCCGCCCGGGGC 5001

Db   1074 GCGGCCGCCGNGNNGGCGCGCGCCGGGGCGGGNGCCCGCGGCGGCNCGCCCGC 1133
              | ||| | |||| | |||||| || |||| || ||||| ||
Qy   5002 CGCTCGTGCAGCCGCGTACCGTCCGGTTCGGCGGTGTCGGTCCGCGCAGTCCGGAG 5061

Db   1134 GCGCCCNCGNGNGCGCGCNGCGTACCGCCGCGCGCGNNNCNCGCCGCACGCCCGCNCG 1193
              |||| | || ||| || |  | |||||||| || |
Qy   5062 ACGAAGATGATGCCCGGACACCCGCACACCTCCGGGTTCCACGAGCATCTCCGTGAGTTGCTCTCCTGGTC 5121

Db   1194 CCGGCGCGGCGGAGCCCGCGGCCGCGGCGCGGCGGCCCGCACNGACGCGGGAGNCGCGGNC 1253
              || ||| | || | | |||| ||| || | || |
Qy   5122 GAGCCGCGGCGGCGGGCGTTCGGCTGGCGAGCACGGCCGCCGCGCCCGGCGGGTCTCAGCGGCCTGC 5181

Db   1254 CGGCCGCGGCACCGCGGCGCGGCGGCGCCGGCGCGGCGGCGGNNC 1313
              | ||| |||||||| ||||| || |||| || ||| || ||||| |
Qy   5182 CCCATCACCTGGGCCGGCGCCGCGCCCGGGGAAGAAGGGTTGTCCAGTTCCGGGGGTGACCAGTCCG 5241

Db   1314 NNCNNCGACGGCGNCGCGGCCGCGGCCGCGCGCGGCGCCGCGGCGGCGNGCCGGAC 1373
              || |||| | | |||| |  |||
Qy   5242 ACCAGCTCGGCGCGGCGGCGCGCGGCTGCTCTAGCCCAGGCCGCGTCCAGTGCGGTC 5301

Db   1374 GCGCCGCGCGCAGCCGAGTCGCGGGTCGGCGGCGGCGCCCNNCCGCGCNNCCCGGGCC 1433
              ||||  ||| |||||| || |||| | |||||
Qy   5302 AGCACCCGAGTCGCGGGTGCCGGGTGCCCGGTGCCACCGCGCGC 5339

Db   1434 TACAACATCGGCGCCGCCCACAACGAGCAGAAGAAC 1471

RESULT 11
LOCUS       AW064148       273 bp    mRNA       linear   EST 07-DEC-2000
DEFINITION  SP0582 KRIBB Human CD4 intrathymic T-cell cDNA library Homo sapiens
            cDNA 3', mRNA sequence.
ACCESSION   AW064148
VERSION     AW064148.1   GI:8888085
KEYWORDS    EST.
SOURCE      human.

REFERENCE   Chlamydomonadaceae; Chlamydomonas.
            1 (bases 1 to 1281)
AUTHORS     Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE       Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular plants; project phase 2
JOURNAL     Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.

FEATURES           Location/Qualifiers
     source        1..1281
                   /organism="Chlamydomonas reinhardtii"
                   /strain="CC-1690 wild type mt+ 21gr"
                   /db_xref="taxon:3055"
                   /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                   II"
                   /note="vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                   XhoI; This library, constructed by John Davies and Jeffrey
                   McDermott, combines cDNAs from CC-1690 cells grown to
                   mid-log phase in TAP (acetate-containing) medium in the
                   light, TAP medium in the dark, HS (minimal) medium in
                   ambient levels of CO2 and HS medium bubbled with 5% CO2
                   PolyA mRNA was purified from each sample, pooled and cDNA
                   synthesized. The cDNA was directionally cloned into lambda
                   ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                   pBluescript II SK- plasmids were excised from the lambda
                   ZAP clones by superinfection with ExAssist (Stratagene)
                   phage. The library was normalized using method 4 described
                   in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT       30 a        697 c       456 g       15 t       83 others
ORIGIN

Query Match       1.6%;     Score 106.8;   DB 12;   Length 1281;
Best Local Similarity  43.8%;   Pred. No. 1.1e-06;
Matches  538;   Conservative    0;   Mismatches  678;   Indels   13;   Gaps    6;

Qy     3145 GGGAGGACTGCCCGGTCGAGGACAGCCCCATCGAGCTACTCCGTTCCTGCGTGCGGCGCGCC 3204
            ||| |||| || || || || ||   |  | |  |   |||  || | ||||| |   |
Db       40 GGGNGGACCCNNNGNCCGGNGCCCGGGCGCGGCCCCCGCCCNCGNNCCGCGGCCCCC 99

Qy     3205 AGTCGAGGGGCGTCCCCGCATCGAGGCGTCCCTCATCGGCCGGCGGCCGTCGTCGGCC 3264
            |  |   | || || ||||  || ||  || | | |  |||  ||| || ||| |  |
Db      100 CCGGGGCGCCGGCCCCGGCCCCGGCGCCCCCGC--CCCCCCGGCCCCCCGCCCCCG 157

Qy     3265 CGGCCCCCCGTCTCCCCGCAGGCTCACCGACTGTGATCGGCGGACCACAGCCAAGGTGTATC 3324
            ||| ||||   |   ||| ||||| |||| | |  || |||  |||||||| || | ||
Db      158 CGCCCCCCGGCCCCGGGCGGGGGCGGCGGCGCNNGCCCGCGGGNGGCGGCGGNNNCC 217

Qy     3325 TCACCCCATGACCACGACCATCCTCGTTCACCGCGGGCGGAGCGGGGCTTCATTCCTCCGCCCTA 3384
            |  || || || | ||  ||  |  ||  ||||| | ||| | | || |  ||  || ||||||
Db      218 GGCCNGCCGGNCCGCGGGGGNNNNCGGCCCGGCGGCCCCCCGGCCGTCGCGGGTGACCGTCCTCGACAA 277

Qy     3385 CGTTCCGCCGGCTCCTGTCCGCCGGGCCCCGGGGCCCCCCCNNCCGCGCCGCC 3444
            |||| |  | ||  | |  ||  ||||  || | ||  ||||||  |  | ||
Db      278 CCTCCTTTGTCGCNNNNNNNGGGCGCNNNCC-CCNNCCCCCCCCCCGCGCNCNC 336

Qy     3445 ACTCACCTACGCCCGGCAGCCTCGCCCCCGCCTGCACGGCGGTGCGTGACCATCCCGGCCTCAC 3504
            | |  |||||||  ||  |||| ||  |||  ||  || | || | | |||||||  || ||
Db      337 CGCCCCCCGCCCCNCCCCCCGCCCCGGGCCGGCNCGGCGCNCGCNCGCCCCCC 396

Qy     3505 CTTCGTCCAGGGCGACGTCTGCGACACCGACACGGCGGCCGGCGGCAGCTGGCCGCGGCACGA 3564
            |  || || |  |||| |  ||| ||||||| |   || ||| || ||| ||  ||   || |
Db      397 CCGCGCGCGGNNGGGGGGGGCGGCGCNCGCCCCGGCGCCCGGCGGC--GCGCCCCGGCGGCNC 454

Qy     3565 CGACATCGTCACTTCGCCGCCGAGTCGCACGTCGGCGCCGCCGCGGCCGCCGGCCCCNGCNCCCC 3624
            ||| | | || ||| || ||  |    || ||| |  |||  || ||||||  ||  |  || |
Db      455 GCCCCNCGCCGGCGCCGCCCGCNNCCGCCCGGCCGCCGCCCCGGCGCGCCCCNGCGNCGCCGGC 514

Qy     3625 CTTCACCCGCCACCAAGCTGCTGGGCACCCAGGTCCTCGTCGACGCCCGCCTCGACGCCCGCACGG 3684
            | ||| ||  || || || ||  || |  ||   |    || |||  | |   |  |   ||| ||
Db      515 CGGCGGCCCGCCCCCCNGCCCCGGCCCCCCCGGCGCCCCCCGGGGCCCNCGGGCCCNCGGGCG 574

Qy     3685 TGTGCGCACCTTCGTCGACGTCTCCACCGACGAGGTGTACGGCTCCTCCCAGCGGCGGGNCC 3744
            |    |  ||| || || |  | ||  ||   ||| || | |  |   |   ||  || |
Db      575 CGGCGGCCCCCGNGGCCGGCNNCCCGGCCCNNCCCCGCNGGCGGCGGCNGCGCGCGGNCC 634

Qy     3745 CGCCCGCGGAGAGC---GACCCCTGCTTCCGACCTCGGCGGCGGCGGCCGTCGAAGGGCGGC 3801
            | || ||| || |  | |||  ||   | | | || ||||| ||  ||  |     |||||
Db      635 CCCCCCGCCGGGCGGCGGNCACCCGGGGNNCACCCACGGCCTGGACGTCCGGTGACCCG 694

Qy     3802 CTCGGACCTCATGGCGCTCGCCCACCACCACGGCCTGCGGCCTGGGTCCGGTGACCCG 3861
            | ||    ||    |||  ||  |  |||||   |    || ||     ||  | |||
Db      695 CCNCGCCCGCCCCGCCCCCCGCCCCNGCCCCGCGGCACGGCGGGGCCCCCGGGCGCNC 754

Qy     3862 CTGTTCGAACAACTTCGGCCCACCAGCATCCCAGAGAGCTCATACCCGCGTTCCTGAC 3921
            ||  | | ||| ||  ||   ||  || | |||  |  || | |||||||| |||||
Db      755 CCGGGGCGCNCCCGCGGGGCGGGGCGCGGCGGCGGGCGGGCGGGGCGGGGACT 814

Qy     3922 CAGCCTC-CTGTCCGGCGGCACCGTTCCCTCTACGGCGCAGCGGCGGCCACGTGCGCGGACACT 3980
            | |||| |  ||| || || |||  || | | | |  | | || ||  || ||  ||| ||
Db      815 GCGCCNCGGCGGGCCCCCCCGCCNCGCGCCCGCCCGGCGCCGCCCGGGGCCCGGGCCCCG 874

Qy     3981 GGCTGCACGTGACGACCACGTCAGGGCCGTCGAACTCGTCGGCGCGCGTCGGGCCCGCCCGG 4040
            ||    |  |   ||   | | |  |  ||   |  ||     |   || |  |  || |
Db      875 GGCGCCCGGGCCCGGCCGGGGNGCGGGGCGGGCGGGNGGCCCCGCGGC 934

Qy     4041 GAGAGATCTACAAACATCGGGGCGGCAGCTGCGTCGCTGCCCAACCTGCCCGCGCCACCGGT 4100
            ||  |   || | |   |  | | |   |  ||||||| | | |  || | || || || ||
Db      935 GCCCGCGCGGCNCCCCCNCGGGCCCGCGGCNCCCCGCGCCGGCGGGNGCGCGCGGCGGGC 994

Qy     4101 TGCT----CGCACTGTGCGGCCCGCGCCCGGGGGCGGCCCGCCCNCCGCCTGCTTGACCG 4156
                    |||  |  |  || ||||| || ||| |   || |   |   |  || ||
Db      995 CNCCGGCGCCGCGGGGGGCCCGCGCGCGCGGGGGGGGGGGCGCCCGCGCGCGGGG 1054

Qy     4157 GGGCACGACCAGGCGCTACGGCGGTGCCCCCCCGCCGCCCGCGGGAAACTCGGTTACCGG 4216
            ||||||  |   |  |     ||||| ||  ||| |  | | |  || | |  || ||||
Db     1055 GGCCGCGGNCCCNNGGGGGGCGCGGCGGCCCGGGCCCCGCCCCGGGCCGCCGGG 1114

Qy     4217 CCGCGCCACCGACTTCGGCAGGTCGGCGGCCGAGAGCGGCACGAG 4276
            | || ||  ||| ||||| | | | || ||  || ||| |||||
Db     1115 CGCCCGGCCCNCGGGCCCGGCGGGGCGCGGCCCGGCGCCGCGGCCGGGCGGCGGCGCGCGCCC 1174

Qy     4277 GACTGGTGGCGTCCCCTGCTGCCGACATGACGTCGGGCGCGGACCGCAACCACCCGGCC 4336
            || || | | ||  ||| ||  ||  | |  |  | | |  || || ||| || ||| |
Db     1175 GGGCGGCGGCCCGCGCGCCGGCGCNCCGCGGCCCGCGCNNCCCCCGCCGGGCCCCCCCCC 1234

Qy     4337 CCGGCCGGCCACACCGGCCGGCCGGCCGG 4365
            ||||  ||||| ||  ||  ||  | |||
Db     1235 CCGGCCCGGCGGCGGCGGCGGCGGCCCAG 1263

RESULT 13
BH606535
LOCUS       BH606535    BOHAT59TR BOHA Brassica oleracea genomic clone BOHAT59, DNA
                        828 bp    DNA        linear    GSS  15-DEC-2001
DEFINITION  BOHAT59TR BOHA Brassica oleracea genomic clone BOHAT59, DNA
            sequence.
ACCESSION   BH606535
VERSION     BH606535.1    GI:17858981
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 828)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other GSSs: BOHAT59TF
            Contact: Chris Town

```
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA.
                Tel: 301-838-3523
                Fax: 301-838-0208
                Email: cdtown@tigr.org
                DNA is from a doubled haploid provided by Tom Osborn.
                Seq primer: TR
                Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHAT59"
                     /clone_lib="BOHA"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT    217 a    215 c    185 g    211 t
ORIGIN

Query Match             1.5%;    Score 104.6;  DB 17;  Length 828;
Best Local Similarity  50.9%;    Pred. No. 2.4e-06;
Matches   366;   Conservative   0;   Mismatches 329;   Indels   24;   Gaps   4;

Qy  3332 ATGACCACGACCATCCTCGTCACCGGCCGGAGCGGGCTTCATTCGCTCCGCTACGTCCGC 3391
         | ||||||| || ||||||| | ||||| ||| | ||| |||| | ||| | |||
Db    63 AAGCCCAAGAGAACATCCTCATAACCGGAGCCCCGGGTTTCATCGCCCTCCCAGTCGCCAAC 122

Qy  3392 CGGCTCCTGTCGCCCGGGGCCCCCGGCGTCGGCGTGACCGTCGGCGGTGACCGTCGGCGAC 3451
         ||| ||| | | |||||  |||||||| | |||  || || |||||||| || || ||||
Db   123 AGACTAGTCCGCACCTACCCTCACTACAAAATCGTG------GTCCTCGACAAGCTCGAC 176

Qy  3452 TACGGCCGGCAGCCTCCCCCGCCTGCACGCGGTGCGTGACCATCCCGGCCTCACCTTCGTC 3511
         ||| | ||  | ||| | ||| || |||||| || ||  ||||||| ||| |||||||||
Db   177 TACTGCTCCAACCTCAAAAACCTTAACCCTTCGTCGTCGACACCGTCGG------CCGGCGGCAGGAC 236

Qy  3512 CAGGGCGACGTGTGGACACCGGTCGTCGACACGCTGG------CCGGCGGCACGAC 3565
         ||||||||| || | || ||| | |||||||||| |      |||| ||| ||||
Db   237 AAGGGCGACATCGCCAGCGACGACGACCTCGTCAGCTACCTTCGTCATCACCGAGAACATCGAC 296

Qy  3566 GACATCGCTGCACTTCGCGGCCGGCCGAGTCGCACGTCGACCGCTCCATCACCGACGCGGTGCC 3625
         ||| | |   || |||||  | | ||||  || || || ||||| |||| ||| |  |
Db   297 ACCATCATGCCACTTCGCCGTCAGACCCGCGTCGAGCCACGTGCGCGAACTCCTTCGGCAACAGCTTCGAG 356

Qy  3626 TTCACCCGCCACCAAACGTGCTGGCACCCAGGTCCTCGCGACGC---CGCGCTCCGCCAC 3682
         |||||  | || ||||  || |  ||| |||| || |  |||     |||| ||||| ||
Db   357 TTCACCAAGAACAACATCTACGCGCACGCACGTCCTTTTAGAAGCTTGCAAAGTCACTGGA 416

Qy  3683 GGTGTGCGCACCTTCGTCGACGTCTTCCACCGACGAGTGTACGGCTTCCCTCCCGCACGGG 3742
         |||| || |||| ||| || | | | ||| |||||||| | || | || | ||||| |||
Db   417 CAGATCAGGAGGTTTATCCACGTAGACAGTGAGTCTACGGAGAGACTGATGAGGAT 476

Qy  3743 GCCGGCCGGCGGAGGCGA-------CCCCTGCTTCCGGCCGTTCGGCGTACGGCGGCGTCG 3793
         ||| | ||  | || |         ||| || || ||  | || | || |||||| ||||
Db   477 GCTTCCGTTGGGAACCATGAGAGCTTCTCAGCTGCTTCCGACTAATCCTTACTCCGGTACT 536

Qy  3794 AAGGGCGGCCTCGGACCTCATCGCGGCTCGCCCACCACCGACCTGGACGTCCGG 3853
         |||| || |||| |||| || || |||||  |||| | |||| || |||||||
Db   537 AAAGCTTGGTGCTGAGATGTTGTGATGGCGTATGGGAGGATCGTATGGTTACCGGTTATA 596

Qy  3854 GTGACCCGCTGTTCGAACAACTTCGGCCCCCCACCAGCATCCCGAGAAGCTCATACCGCGC 3913
         ||| | || || | | || || |||  ||| |||| ||| || ||| || || | || |
Db   597 ACCACGCGTGGGAATAATGTTTATGGTCCGAACCAGTTCCTGAGAGTTGATTCCTAAG 656

Qy  3914 TTCCTACCAGCCTGCTGTCTCCTGTCCGCGGCCACCGTTCCCCTCTACGGCGGCGGCCACGTG 3973
         ||| |   |||| | |||  | |||| ||  | ||||  ||  | ||||||||| |||
Db   657 TTCATGTTGTTGGCTATGAATGGGAGCCGCTTCCATCACCGGGGATGGGATGGATTCTAATGTA 716

Qy  3974 CGGGACTGGCTGACGCTCGACGACCTCGAACTCGTCCGGTCCGG 4032
         ||  | |||| ||| || |  |  ||| |||| || |||
Db   717 AGGAGTTACTTATACTGTGGAAGACGTTGCTGAGACGTGCGGAGGCCTTTGCTTTGAGTTGTTCTGCACAAAGGG 775
```

```
LOCUS       BG809816                1651 bp    mRNA        linear    EST  22-MAY-2001
DEFINITION  mgct001xk18f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct001xk18f 5', mRNA sequence.
ACCESSION   BG809816
VERSION     BG809816.1  GI:14180796
KEYWORDS    EST.
SOURCE      Magnaporthe grisea.
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES             Location/Qualifiers
     source          1..1651
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct001xk18f"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /dev_stage="Germinated conidia on appressorium-inductive
                     surface"
                     /note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed from conidia germinated for 5-8 hr
                     on an inductive surface. The library contains over 55,000
                     clones with average insert size of 1.5 kbp."
BASE COUNT    66 a    320 c    769 g    21 t    475 others
ORIGIN

Query Match             1.5%;    Score 104.6;  DB 12;  Length 1651;
Best Local Similarity  33.3%;    Pred. No. 2.5e-06;
Matches  425;   Conservative   0;   Mismatches 850;   Indels   0;   Gaps   3;

Qy  2601 CAGGGCGTCACCGACTTCGCCCCCAATCGCCCGATCCCGGCGGCCCGGCCCGGCTGCTG 2660
         | ||||| | ||  |   ||| |||   | |||||| ||||||||||||||||||| ||
Db  1539 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCCCCCGGCGGCGGCGGCGGCGGCGGCGG-GG 1481

Qy  2661 CTCACCCCGGTCGCGGACCCGTCCGCCCTTCGGCGTCGGCGAGGTGACCGGGACGGGAAC 2720
         | || |||||| | |||||  | ||| | ||||  | || ||||   ||||||  ||
Db  1480 CGGCGGCGGGCGGGCGGGCCCGGCNGGGCGGGCGGGGGGCCCCGGCGGGGCGGGGCGGGC 1421

Qy  2721 GTGCTGCCTTGGAGGAGAAACCCGACGTCCCGGCGGCGGGGCCCCGGCCGCTCATCGGCGTG 2780
         | || | | | | || | |  ||| | ||||||||||||| |||||||||||| ||||||
Db  1420 CGCCCCGGCCGCGGCGGGGCGGCCCGGCCCGGCCCCGGGCCCCGGCCCCCGCGCGGGCGG 1361

Qy  2781 TACGGCCTTCAGCCGCGGCCGCCGGGCCGTCCACGAGGCCGTACGGGCCCATCACCCGGGC 2840
         ||| || | | |||||||||||| | |||| || || |||||||  | || ||| |||||
Db  1360 GCCCCCCCGGCGGCGCGGCGGCGGGGCGGGGCCCCCCCCCCGGCGGGGGCGGGGCGGGGCC 1301

Qy  2841 GAGCTGGAATCACCACCGCGTCAGTGATGATCGAGAAGCTCATACGGGCGGGCCTACGGGGC 2900
         ||| | || | | ||| || | |||  || |  || |  |  || || || || |||||||
Db  1300 CGGCGGCGGCGCGGCGNGCGNGCCC-CTGGCGCGCAGACCGGCGGGCCGGCGGGGCGCCC 1241

Qy  2901 GAGACCACCACCCGGCC-CTGGCGCGACACCGGCGGGGCGGCGGGGCCCGCGGGGGCCGCGC 2959
         ||| | | |||  |||   |||| |||| | || ||||| || || |||||| || ||
Db  1240 GCGNGCCGGCGGGCGGGCCGGNNGGCCGNGGCGCGCGGGGCGGCGCGGGGGCGGCCCGC 1181

Qy  2960 CCGTCACGTCCTGACGGACTGAGGGCCGCCGCCGGGGGCGCGGGGGAAGGTGCACAGCAC 3019
         | ||| |||| ||| | |||  ||  | || || |||||| || ||| ||| | |||||
Db  1180 GCGGGGCGCGGCGGCCGGCGGCGGGGCGCCCGGCCGNGCGCGGGGGCGGNGCGGCCCCGCG 1121
```

```
REFERENCE    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
             1
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        BAC end sequences of Library PTB
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1542)
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
             and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
             1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jP/,
             Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library PTB This BAC end
             was generated from the chimpanzee BAC library PTB This BAC end
             clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES     Location/Qualifiers
     source  1..1542
             /organism="Pan troglodytes"
             /db_xref="taxon:9598"
             /clone="PTB-007N10.F"
             /sex="male"
             /cell_type="lymphoblast"
             /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT    61 a    657 c    614 g    19 t    191 others
ORIGIN
```

```
Query Match          1.5%;   Score 104.2;   DB 17;   Length 1542;
Best Local Similarity  43.3%;   Pred. No. 2.8e-06;
Matches 544;  Conservative 0;  Mismatches 694;  Indels 19;  Gaps 5;
```

```
Qy  729 CCGCTCGACGGCCGTCTCCAGCAGGCGCAGGCCCTCGCTCGAGCAGGCGCCGGCCTCGTCGAGGGT 788
Db 1500 CGGCTGNCCGCGCCGGCGGACAGACGCGACGGCGCGGCGCCGGNGGCGCCGGCCCAGTCGCGCGC 1441
Qy  789 GAACGGCGGTGCCAGCCGGAGGATGTGGGCGCCCAGGGAGGTGCGNCGCCCCAGGTCGAG 848
Db 1440 CGGGGCGGCCGCGCCCCNGCGCGACGGCCCNGCGNGGCGCGNGCGCGGCGNGNGCGCGCCCG 1381
Qy  849 GGCGGTGGTGTAGACGGCCCGGCTCTCGGGGCGGGTGCCCG-GCCGACGGCGTCGG 907
Db 1380 GNCNCGGCGCCGGCGCGGGCCGCGCGCGCGCCCGCGGCGTGTGGCNCGCGCGG 1321
Qy  908 TGAGGAACTCCAGGCCCCACAGCAGTCCGAGGCCGCGTACCTGCCGAGCTGGGGGAAGC 967
Db 1320 CGGGCGCGGCCGCGCCCGCTCGCCCACCACCGCGCGGGNGGCGGCGCGGCGGGGGC 1261
Qy  968 GGGACTCCAGGCGCGCCAGCCGGCGCGCTCCTGGATGAGCTGGCCGGAGGACGGCCACGCGGTGA 1027
Db 1260 GCGGCGCNCNGTGGCCGCGCGCGGCGCGCCCCGCGCGCGCGCGGGCGATCCCGCGGCGGC 1201
Qy 1028 TCAGCGGTGGCGCGCTGGACGACCTTCCAGCGTGGCGCGCGGCGGGGCGGATCCCCAGTGGGT 1087
Db 1200 GCCCGCGCGGCGCGGCGGCGCGGCCGCNCGGC---CNCGCGCGCGGCGCCCGCCGCAGCTTCCG 1144
Qy 1088 TGCTCGCGTACGTCGAGCGTACGCGGGTGGCCGCGCCCCGGGGTGCGGGGCGCAGCTTCCG 1147
Db 1143 CCCGCGCGCNCGGCCGGGCGCNCCGCNCCGCCCCGCGCGCGCGCCGCCCCCGNCGGGGCGGC 1084
Qy 1148 CGCGTCGGGCCAGCACGCGCGAAGGGAATCCGCTCGCGGTGCCTTGGACACGCATCGCCA 1207
Db 1083 GCGGGCGCGCGCGGCGCGGGGCGGCCCGNGCGCCNGCGCGCGCGCGCCGGCGCGGCGGCC 1024
Qy 1208 GGTCCG---GCTCGATGCCGAACAGTTCGCTGGCGAGGAAGGCGCGGGTGGTCGGGCGGCGC 1264
```

```
Qy 3020 GCTTGGTCGGCCGGGTCGCGGGTGGCCGAAGGCGGCGATCGTGCGGGGGTCACACGTGTGGG 3079
Db 1120 GCGGCGGGCGGCCGGCGCCCGCCG-GCCCGCCGCGGGGGCGGGCGCGGCCCCGCGCGGCG 1062
Qy 3080 CCCGGTGGTGATCGGCGGGGTGCCGTCGTGTCAGCAACTCCAGTGCCCGTACACCTC 3139
Db 1061 CCGNCGCGNGGCNCCGCNCCGCGNCCCNCNCNNNCCCNCNNNNNCCCNN 1002
Qy 3140 CATCGGGAGGACGTCGCGGGTCGAGGACAGCGCCATCGAGTACTCCGTCCTCGCTCCGCGG 3199
Db 1001 NNNCCNNNCCCCCNNNCCCCCNCNNNNCCCNNCNNNNNCNNNNCNCN 942
Qy 3200 CGCCCAGTCGAGGGGGCGTCCCTCATCGAGGCGTCCCNNNCCCNNNCCCNNNCCCNNCCN 3259
Db 941 NNCNNNNCCCNNNNNCNCNNNCNCNNNNNGCGGACCACCAGCAAGGT 882
Qy 3260 CGGCCCGGCCCCCGTCTCCCGAGGCCTCCGACGACCTGTGATCGGCGACCACCAGCAAGGT 3319
Db 881 CCCCCCCNNNNCNCNNNNCCCCCCCNNNCNNNNCNCNCNNCCCCCCCNNNC 822
Qy 3320 GTATCTCACCCCATGACCACGACCATCCTCGTCACCGGCGGAGCGGGCTTCATTCGCTCC 3379
Db 821 CCNNCCNCCCCCCCCCNNNCCCNNNCCCCNCCCCNCNNNNCCCCCCCCNNNC 762
Qy 3380 GCCTACGTCCGGGCTCCTCGTCTCCCCCGGGGCCCCGGGGCCGGTCGGCGGTGACCGTCCTC 3439
Db 761 NCCNCNNNNCCNNNCNCNCCCCCCNNNCCCCCNCCCNNNCCCNCNCNCNCC 702
Qy 3440 GACAAACTCACCTACGCCGGCCAGCCTCGCCCGCCTGCACGCGGTGCGTGACCATCCCGGC 3499
Db 701 CCNCNCCCCNNCCCCCCNCNCNCNNNCCCCNCCCNCNCNCNCNNCNNNC 642
Qy 3500 CTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTCGTCGACACGCTGGCCGCGCCGG 3559
Db 641 CNCCNCNCNCNNCNNNCCCCNNNCNCNNNNCNCNNNCNNNNCNNNNCNNN 582
Qy 3560 CACGACGACATCGTGCACTTCGGCGGCGACTTCGGCCGAGTCGCACGTCGAGCGTCGACGACGC 3619
Db 581 CCNCNCCCCNCCCNCNNCNCCCCNCNCCCCNNCNCNCCCNCCCNCCCCNNN 522
Qy 3620 GGTGCCTTCACCCGACCAACGTGCTGGGCACCCAGGTCCTGGCACGCGCGGCCGCTCCCGC 3679
Db 521 NNCCCNNNCCCCCNNCCNCNCNNCNNNCNNCNCNCNNCCCCNNNNNNC 462
Qy 3680 CACGCGTGTGCGCCACCTTCGTTGCACGTCTCCACCGACGAGGTGTACGGCTTCCCGCAC 3739
Db 461 CNNNCNNCNCCCCCCCCNNNCNCNCNCNCNNCNCNCNCNNNCCCNCNC 402
Qy 3740 GGGGGCCGCGCGGAGAGGACGACCCCCTGCTTCCGACCTCGCCGTCGAGGCG 3799
Db 401 NCNNNCCNNNNNCCNCNCNCNCNNNCCCNCNCNCNNNNNNCCNNAAAAANNN 342
Qy 3800 GCCTCGGACCTCATGCGCCTCGCCCACCACCGCACCGCTGGACGTTCCGGGTGACC 3859
Db 341 CTCTCTCCCNNCNCNNNNNCCCNNNCNCNCNCNCNNNNNNNNCNNCNCNNCNCNNNCN 282
Qy 3860 CGCTGTTCGAACAACTTC 3877
Db 281 NCNNNNCNNNNNCNNNNNC 264
```

RESULT 15

```
AG032943/c
LOCUS       AG032943           1542 bp    DNA       linear    GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-007N10.F, genomic survey sequence.
ACCESSION   AG032943
VERSION     AG032943.1  GI:16559816
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-007N10.F.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Db  1023  GGGCGNGGCGCCCGCCGCCCGNCGCGGCGGCGCCGGCCGCGCCGCGNCGCGCCCCGCGCG  964
Qy  1265  CGGTGAGGACCTCGTCGGCGACGAGCAGCACGCCCGCGTCCGGCCAGGCGCCGGCGATCC  1324
                - | - | - | |||||| - | - ||| - | | - | ||| - | ||| -  |||
Db   963  GCCGGCGCGCGNCCGCGGCGGCGGCGGGGGCGGCGGCGGCGGNGGCGGCGGCGGGCGGC   904
Qy  1325  GCTCCCAGTAGCCGGGGGGCGGCACGATGACGCCTGCCGCGCCGAGGACGGGTTCGAAGA  1384
                -| - |||||||||| - | - ||| - | | - ||| - | |||
Db   903  GGGCGCCCCGCCGGGGGGGGGNGCGCCCGGGCGGCGCCGCGCCCGGCGGCGGCGC   844
Qy  1385  CCAGGGCCGAGACGTTGGGCTTCTCCGCGATGTGCCGGCGGCACGAGGGTCCGCACCGCA  1444
                 | - || - |||| - | | - ||| - | - | |||| - || - |||| - |||
Db   843  GGGGGCGGCGGCGNNCGCGGCGCGGCGGCGGCGCGGGGCGCGGCGGCGCGGCGCNCNCCGCG  784
Qy  1445  CGTCCGCACGAGGGGTACTCCAGGCCCAGGGACACGCCGTAGCCAGTAGGGCTGTAGCCA  1504
                || - || - ||| | - | | - | ||| | - || - | | - ||| -
Db   783  GCGNGCCGNNNCGGCCGCGCCGCGCGGGGGGGGGGGGCGGCGGGCGCGGGG  724
Qy  1505  GCACGCTGTTGCCGCTGAAGGCCTCGTGCCGATGTCCCAGTGGACCAGCATCCGGGCGC  1564
                || - |||            | | - ||| | | - | |||
Db   723  GCCGCCGCGGCCCGCCGGCGGGNGCGGCC --------- GGGCGCCGCGCCCGCGCC  675
Qy  1565  CCATGGTCTTGCCGTGGAAGCCGTGGCGCAGGGCGCAGATCCGGTTGCGGCCCGGCGCGG  1624
                || - | - - ||| - | - | ||| | ||| - |||
Db   674  GCGGCCGCGCCGGCGCCGGCGGGGGCGCGCCCCCGCGCCCNCCGCGCCCCCGCGCCGCGN  615
Qy  1625  CGGTCGCCTGGACGACCCGCAGGGCGCGCCTCGACCACCTCGCGGCGGTGGAGAAGAAGG  1684
                | | - |||| ||| - | | - | - | - | - | | - |
Db   614  CCGCGTGNGCGCGGCGCCGCGCCCCGCGCCCGCGCGCGCGNNGGCGCCGGCGGGCCG  555
Qy  1685  CGTAGGTGTCGAGCTGTTCGGGCAGCAGCCTGGCGAGCAGTTC --CAGCAGGCCGGCGCGCGG  1743
                || || - | - | - | - | |||| - ||| - ||| |||| ||||
Db   554  CGGGGGGGCNCGCGGCGGGCGGCGGCNNGCCCGCGCGGGGCNCNCCCCGCGCCCC  495
Qy  1744  TCCGGCCGGTGGCGCTGCGTGTCGTGGACGTTCCACAGGCCGGCCTGGGTGGTTGAGTGCCTCG  1803
                || - | - | ||| - ||| - - | |||| - | |||
Db   494  NCCCNNCCCCGCCGCCGGCCCGGCNCNNNGGCCGGGCNNNAAAGAA  435
Qy  1804  ACGACCTCCGGTGCCCGTGCCCAGTGACTGGGTGAGGTCCCGGCCGCGGAAGTCGAGG  1863
                | - | - | - | - | | - ||| - | |||| - |
Db   434  TTGGACGGAATATTTTNTTTTTTCTTTGTNTNFTGTCGGCGGCGCCCCGCCGNGG  375
Qy  1864  TACTTGGTTGCCGTCCAGGTCGGTCAGAACGGACCCGCGTCCCTCCGCGAAGACCCGGCGT  1923
                | | - | - | - | - | - | - | - | - | ||
Db   374  GGGGGGNGCGGGGNCGCGGCGCCGCGCCGCGCCCNCNCNCGNCGGCGGGGGGCNGGGGGG  315
Qy  1924  CCGTGGACGGCTTCCTCGGAGGCGCCCGGCGCCCAGGTGGCCGGCCTCCCCGTGCCCAGG  1980
                |||| - | - |||||| - | ||| - | - | ||| | - |||
Db   314  NCNCGCGGCGCGCGGGNCNCGCGGCGGGCGCGGCGGNGCGCCGCCCGCCGG  258

Search completed: May 28, 2003, 21:04:52
Job time : 5799 secs
```

```
                    GenCore version 5.1.4_p5_4578
          Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       May 28, 2003, 16:02:59 ; Search time 141 Seconds
                                       (without alignments)
                                       14907.536 Million cell updates/sec

Title:          US-09-922-683-7
Perfect score:  6854
Sequence:       1 CTGCAGGGTTCCCTGGTGCA.........TCGGAGCCGATCCGGCTGCAG 6854

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters:        882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_NA:*
            1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

## SUMMARIES

|   | Result No. | Score | % Query Match | Match Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
|   | 1 | 6854 | 100.0 | 6854 | 4 | US-09-194-905-7 | Sequence 7, Appl |
|   | 2 | 529 | 7.7 | 3292 | 3 | US-09-320-878-22 | Sequence 22, Appl |
| c | 3 | 527.4 | 7.7 | 13613 | 4 | US-09-105-537-3 | Sequence 3, Appl |
|   | 4 | 522.6 | 7.6 | 546 | 4 | US-09-194-905-3 | Sequence 3, Appl |
|   | 5 | 497.6 | 7.3 | 2634 | 1 | US-08-196-218-31 | Sequence 31, Appl |
|   | 6 | 497.6 | 7.3 | 2634 | 1 | US-08-681-953-31 | Sequence 31, Appl |
|   | 7 | 465.8 | 6.8 | 2310 | 3 | US-09-036-987A-25 | Sequence 25, Appl |
|   | 8 | 465.8 | 6.8 | 2310 | 4 | US-09-370-700-25 | Sequence 25, Appl |
|   | 9 | 454.2 | 6.6 | 1476 | 4 | US-09-434-288-12 | Sequence 12, Appl |
| c | 10 | 447 | 6.5 | 1014 | 4 | US-09-105-537-13 | Sequence 13, Appl |
| c | 11 | 389.2 | 5.7 | 2219 | 1 | US-08-606-322-1 | Sequence 1, Appl |
| c | 12 | 276.2 | 4.0 | 541 | 4 | US-09-194-905-4 | Sequence 4, Appl |
| c | 13 | 238 | 3.5 | 4403765 | 4 | US-09-103-840A-2 | Sequence 2, Appli |
|   | 14 | 234.8 | 3.4 | 28804 | 2 | US-08-592-874-1 | Sequence 1, Appli |
|   | 15 | 234.8 | 3.4 | 28804 | 3 | US-09-096-942-2 | Sequence 2, Appli |
|   | 16 | 234.8 | 3.4 | 28804 | 3 | US-09-096-867-2 | Sequence 2, Appli |
|   | 17 | 224 | 3.3 | 4776 | 2 | US-08-852-401-1 | Sequence 1, Appli |
| c | 18 | 188.6 | 2.8 | 2639 | 2 | US-07-952-817-8 | Sequence 8, Appli |
| c | 19 | 188.6 | 2.8 | 2639 | 6 | 5210025-1 | Patent No. 5210025 |
| c | 20 | 183.4 | 2.7 | 1272 | 3 | US-09-036-987A-28 | Sequence 28, Appl |
|   | 21 | 183.4 | 2.7 | 1272 | 4 | US-09-370-700-28 | Sequence 28, Appl |
| c | 22 | 159.2 | 2.3 | 5625 | 4 | US-08-961-527-1 | Sequence 1, Appl |
|   | 23 | 157 | 2.3 | 879 | 1 | US-09-105-537-11 | Sequence 11, Appl |
| c | 24 | 153.2 | 2.2 | 9515 | 1 | US-08-920-812-13 | Sequence 13, Appl |
| c | 25 | 153.2 | 2.2 | 9515 | 1 | US-08-920-827-13 | Sequence 13, Appl |
| c | 26 | 153.2 | 2.2 | 9515 | 1 | US-08-921-177-13 | Sequence 13, Appl |
| c | 27 | 153.2 | 2.2 | 9515 | 1 | US-08-362-577C-13 | Sequence 13, Appl |
| c | 28 | 153.2 | 2.2 | 9515 | 2 | US-08-920-828-13 | Sequence 13, Appl |
| c | 29 | 150.6 | 2.2 | 4403765 | 4 | US-09-103-840A-2 | Sequence 2, Appl |
|   | 30 | 132.6 | 1.9 | 44377 | 2 | US-08-804-227C-7 | Sequence 7, Appl |
|   | 31 | 132.6 | 1.9 | 44377 | 2 | US-08-804-198-1 | Sequence 1, Appl |
| c | 32 | 130.8 | 1.9 | 8438 | 1 | US-07-945-283-1 | Sequence 1, Appl |
|   | 33 | 127 | 1.9 | 4411529 | 4 | US-09-103-840A-1 | Sequence 1, Appli |
| c | 34 | 125.4 | 1.8 | 4257 | 2 | US-08-690-473-1 | Sequence 1, Appl |
| c | 35 | 125.4 | 1.8 | 4257 | 4 | US-09-259-821A-1 | Sequence 1, Appl |
| c | 36 | 125 | 1.8 | 4257 | 4 | US-08-83-659-1 | Sequence 1, Appl |
| c | 37 | 125 | 1.8 | 1064 | 4 | US-08-858-207A-88 | Sequence 88, Appl |
|   | 38 | 122 | 1.7 | 15872 | 4 | US-09-105-537-1 | Sequence 1, Appl |
|   | 39 | 118.6 | 1.7 | 13842 | 4 | US-09-105-537-30 | Sequence 30, Appl |
|   | 40 | 118.6 | 1.7 | 36778 | 4 | US-09-105-537-5 | Sequence 5, Appl |
|   | 41 | 118.6 | 1.7 | 38506 | 3 | US-09-320-878-19 | Sequence 19,.Appl |
| c | 42 | 118.4 | 1.7 | 1281 | 6 | 5221737-1 | Patent No. 5221737 |
| c | 43 | 118.4 | 1.7 | 1608 | 6 | 5221737-6 | Patent No. 5221737 |
| c | 44 | 118 | 1.7 | 4050 | 4 | US-09-543-084A-26 | Sequence 26, Appl |
| c | 45 | 118 | 1.7 | 4093 | 4 | US-09-543-084A-28 | Sequence 28, Appl |

## ALIGNMENTS

```
RESULT 1
US-09-194-905-7
; Sequence 7, Application US/0919905
; Patent No. 6306627
; GENERAL INFORMATION:
;   APPLICANT: DECKER, Heinrich
;   TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
;   TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
;   TITLE OF INVENTION: GLA.O AND THEIR USE
;   NUMBER OF SEQUENCES: 13
;   CORRESPONDENCE ADDRESS:
;     ADDRESSE: FOLEY & LARDNER
;     STREET: 3000 K Street, N.W.
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: U.S.A.
;     ZIP: 20007-5109
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/194,905
;     FILING DATE: 29-JUL-1998
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: WO PCT/EP97/02826
;     FILING DATE: 30-MAY-1997
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: DE 19622783.6
;     FILING DATE: 07-JUN-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granados, Patricia D.
;     REGISTRATION NUMBER: 33,683
;     REFERENCE/DOCKET NUMBER: 026083/0193
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (202) 672-5300
;     TELEFAX: (202) 672-5399
;   INFORMATION FOR SEQ ID NO: 7:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 6854 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
US-09-194-905-7

  Query Match             100.0%;  Score 6854;  DB 4;  Length 6854;
  Best Local Similarity   100.0%;  Pred. No. 0;
```

```
Matches 6854; Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 CTGCAGGGTTCCCTGCTGCACGACCCCGCCCCTGGTCGACGACCAGGGCGCTGTCGCAGAT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTGCAGGGTTCCCTGCTGCACGACCCCGCCCCTGGTCGACGACCAGGGCGCTGTCGCAGAT   60

Qy    61 CGCGGGGATGTCGAGCTGGCTGGTGTCGGTGTGGCACCACGGTGCCCAGTTCCCGGTG    120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CGCGGGGATGTCGAGCTGGCTGGTGTCGGTGTGGCACCACGGTGCCCAGTTCCCGGTG    120

Qy   121 GGCGCGGTTGACCAGCCGGGCGCACCGGCCACCCGCGTCCTTCAGCACCATGTCGAGGCCGATCGTGGG   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGCGCGGTTGACCAGCCGGGCGCACCGGCCACCCGCGTCCTTCAGCACCATGTCGAGGCCGATCGTGGG   180

Qy   181 CTCGTCCCAGACAGCACGCACGGGCGTCGGTCGTGCGCAGGCTGCCGCAGCAGGCTCGGCGCATCGGCGGCGCAT   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTCGTCCCAGACAGCACGCACGGGCGTCGGTCGTGCGCAGGCTGCCGCAGCAGGCTCGGCGCATCGGCGGCGCAT   240

Qy   241 GGCGTGTCGAGCTGAGCGCTGCGCGACGCTGGGGTGGACCCCGAGGGGGTGGACCGCTGCGAGGAG   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGCCTGTCGAGCTGAGCGCTGCGCGACGCTGGGGTGGACCCCGAGGGGGTGGACCGCTGCGAGGAG   300

Qy   301 GTCCCGGAACAGGGCGAGGTTGCGCCCGGTAGACCGGTCGCGGGGATGTCGTAGATCGGGCG   360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTCCCGGAACAGGGCGAGGTTGCGCCCGGTAGACCGGTCGCGGGGATGTCGTAGATCGGGCG   360

Qy   361 CAGGATGCGGAAGGAGTCGGGTACCGGACCAGGGTCCCACCCAGAGCTGCCGCTGGCCGAA   420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAGGATGCGGAAGGAGTCGGGTACCGGACCAGGGTCCCACCCAGAGCTGCCGCTGGCCGAA   420

Qy   421 GACGACGCCGATCGTGCGCCGAGGTGCCGCTGCCGGTAGGGCTCCAGCCCGGCGAC   480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACGACGCCGATCGTGCGCCGAGGTGCCGCTGCCGGTAGGGCTCCAGCCCGGCGAC   480

Qy   481 CGTCAGCGCGCCGAGGTGGGGTCATGATGCCGGTCTTCGTGCGGCGGTATCTCGATCTTCGATCGTTGGTCGACTT   540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CGTGCAGCGGCCGAGGTGGGGGTCATGATGCCGGTCTTCGTGCGGCGGTATCTCGATCTTCGATCGTTGGTCGACTT   540

Qy   541 GCCGGCGTCCGTTGCGCCGATGTAGGCGCGTCTTCGTGCGCGGGGTCAGGAGGTGGAGAGGAGAC   600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCCGGCGTCCGTTGCGCCGATGTAGGCGCGTCTTCGTGCGCGGGGTCAGGAGGTGGAGAGGAGAC   600

Qy   601 GTCGTCGACGCGGCGCGCGCACGACGCGGTACCGCCGGTACCGGCGTCAGGAGGCTGCCGAG   660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GTCGTCGACGCGGCGCGCGCACGACGCGGTACCGCCGGTACCGGCGTCAGGAGGCTGCCGAG   660

Qy   661 CAGGGGCCGGCGTCCGGCTCGCGTTCGGCGAACTCCTTGACGAGGTGTTCGGCCACGATCAC   720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CAGGGGCCGGCGTCCGGCTCGCGTTCGGCGAACTCCTTGACGAGGTGTTCGGCCACGATCAC   720

Qy   721 GCGATCACCCGTCGACGGCGTCGTCCGCAGCGGCCAGGGCCCTCGTCGAGCAGCGCCCTCG   780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GCGATCACCCGTCGACGGCGTCGTCCGCAGCGGCCAGGGCCCTCGTCGAGCAGCGCCCTCG   780

Qy   781 TCGAGGGTGAACGGCGGGTGCCAGCCGCAGGATGTGGCCGCCCAGGAGGTGCGCAGCCCC   840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TCGAGGGTGAACGGCGGGTGCCAGCCGCAGGATGTGGCCGCCCAGGAGGTGCGCAGCCCC   840

Qy   841 AGGTCGAGGCGGTGTGTAGACGGCGGTGGTGTAGACGGCCCGGTCTCGGGGGTGCCCGGCGACG   900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGTCGAGGCGGTGTGTAGACGGCGGTGGTGTAGACGGCCCGGTCTCGGGGGTGCCCGGCGACG   900

Qy   901 GCGTCGGTGACGAACTCCAGGCCCGCCTCCACAGCAGTCCCACAGGCGCGGTACCTGGCCGAGCTGG   960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GCGTCGGTGACGAACTCCAGGCCCGCCCCACAGCAGTCCCACAGGCGCGGTACCTGGCCGAGCTGG   960

Qy   961 GGGAAGCGGGACTCCAGGGCGCCTCAGCGGCGCTCGATGAGTCGCCGGAGGACGCGCACG  1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GGGAAGCGGGACTCCAGGGCGCCTCAGCGGCGCTCGATGAGTCGCCGGAGGACGCGCACG  1020

Qy  1021 CGGTCGATCAGCCGGCTCGCCGCTCGACGCTCCAGCGTGCGGCGGCGATCCCC  1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CGGTCGATCAGCCGGCTCGCCGCTCGACGCTCCAGCGTGCGGCGGCGATCCCC  1080

Qy  1081 AGTGGGTTGCTCGGCTGCGCGGTACGTCGAGGCGTACGCCCTCCGGGGTGGCGGCCTGCCA  1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 AGTGGGTTGCTCGGCTGCGCGGTACGTCGAGGCGTACGCCCTCCGGGGTGGCGGCCTGCCA  1140

Qy  1141 GCTTCCGCGCGTCCGGCCAGCGCCAGACGCGGAGGGGAATCCGCTCGGTGCCTTGGACAGC  1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 GCTTCCGCGCGTCCGGCCAGCGCCAGACGCGGAGGGGAATCCGCTCGGTGCCTTGGACAGC  1200

Qy  1201 ATCGGCCAGTCCGGCTCGATGCCGAACAGTTCGTGGCGGAAGGCGCGGTCGGTGCGCCCG  1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 ATCGGCCAGTCCGGCTCGATGCCGAACAGTTCGTGGCGGAAGGCGCGGTCGGTGCGCCCG  1260

Qy  1261 CCGCCCGGTGAGGAGCTCGTCGGGGACGAGCAGCACGCGCCGCCCGTCCCGGCCAGGCCCGGGCG  1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCGCCCGGTGAGGAGCTCGTCGGGGACGAGCAGCACGCGCCGCCCGTCCCGGCCAGGCCCGGGCG  1320

Qy  1321 ATCCGCTCCCAGTAGCCGGGGGCGGCGGCGCACGATGACCCGCTCGCGCCGGCCGAGGCGGTTCG  1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 ATCCGCTCCCAGTAGCCGGGGGCGGCGGCGCACGATGACCCGCTCGCGCCGGCCGAGGACGGGTTCG  1380

Qy  1381 AAGACCAGGCCGAGACGTTGGGCTTCTCCCGCGATGTGCCGCGCACGAGGGTCGCGCAC  1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 AAGACCAGGCCGAGACGTTGGGCTTCTCCCGCGATGTGCCGCGCACGAGGGTCGCGCAC  1440

Qy  1441 CGCACGTTCGCACGAGGAGTCGGGTACTCCAGGCCCAGGGGCGGTGTAGCCGGTAGGGGCTGTA  1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CGCACGTTCGCACGAGGAGTCGGGTACTCCAGGCCCAGGGGACCAGCCGGTAGGGGCTGTA  1500

Qy  1501 GCCAGCACGCTGTTGCCGCTGAAGGCCTGTTCGGCGCTGGCCGGATGTCCCAGTGGACCATCCGG  1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GCCAGCACGCTGTTGCCGCTGAAGGCCTGTTCGGCGCTGGCCGGATGTCCCAGTGGACCATCCGG  1560

Qy  1561 GCGCCCATGGTCTTGCCGTCGGAAGCCGCCCGGTTGCGGCCCGGC  1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GCGCCCATGGTCTTGCCGTCGGAAGCCGCAGATCCGGTTGCGGCCCGGC  1620

Qy  1621 GCGGCGGTCGCCTGACGACCCCGCAGGGCGCGGCCGCTCGACCACCTCCGGCCGCCGGTGGAGAAG  1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 GCGGCGGTCGCCTGACGACCCCGCAGGGCGCGGCCGCTCGACCACCTCCGGCCGCCGGTGGAGAAG  1680

Qy  1681 AAGGCGTAGGTCGCGAGCGTGTTCGGGCGCTGCGAGCAGTTCCAGCAGGCCCGGCG  1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 AAGGCGTAGGTGTCGAGCTGTTCGGGCGCTGTTCGGGCTGGTTCGAGCAGTTCCAGCAGCCGGCGGTGAGTGCC  1740

Qy  1741 CGGTCCGGCGTGCGCGTCGTCGTCCGCCGGGCCGGCCTGGGTGGTGAGGGTGCC  1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 CGGTCCGGCGTGCGCGTCGTCGTCCGCCGGGCCGGCCTGGGTGGTGAGGGTGCC  1800

Qy  1801 TCGACGACCTCCGGTGCCCAGTGACTGGGCCCCGCGCCGGAGTCCCCGCGCAAGTCG  1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 TCGACGACCTCCGGTGCCCAGTGACTGGGCCCCGCGCCGGAGTCCCCGCGCAAGTCG  1860

Qy  1861 AGGTACTGGTTGCCGTCGAGGTCGGTCAGAAACGGGACCGGTCCTGCGGCGGAAGACCCGG  1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGGTACTGGTTGCCGTCGAGGTCGGTCAGAAACGGGACCGGTCCTGCGGCGGAAGACCCGG  1920

Qy  1921 CGTCCGTGACGGCTTCCTCGGAGGCGGCGGTGCCCGGCCAGTCGCTGCTCGTCGTGACGGCGGAT  1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CGTCCGTGACGGCTTCCTCGGAGGCGGCGGTGCCCGGCCAGTCGCTGCTCGTCGTGACGGCGGAT  1980

Qy  1981 TGCTGTGTCTGCGCTAAGCCTGTCATCGCGCGGCTCGTCGCGGCTCAAACAC  2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 TGCTGTGTCTGCGCTAAGCCTGTCATCGCGCGGCTCGTCGCGGCTCAAACAC  2040

Qy  2041 CGGCCGGGAACTGCGTTGTGCGCCGTTGTGGCCCACCACGGTTGCGGGTCTGTTCGAAAGTTGTTCGGACGCGGCCC  2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGGCCGGGAACTGCGTTGTGCGCCGTTGTGGCCCACCACGGTTGCGGGTCTGTTCGAAAGTTGTTCGGACGCGGCCC  2100

Qy  2101 TTGAACACACCCGTCGCAAGAGTTTGCGGGTTGTTTCAGAAAGTTGTTCGGAGCGGCCC  2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 TTGAACACACCCGTCGCAAGAGTTTGCGGGTTGTTTCAGAAAGTTGTTCGGAGCGGCCC  2160
```

```
Qy  2161  CGGCACTCTGGTTGAGTCGACGTGCTTACGGCGCCACCACGCCTCACGTTCGAGGAGGGA  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  CGGCACTCTGGTTGAGTCGACGTGCTTACGGCGCCACCACGCCTCACGTTCGAGGAGGGA  2220

Qy  2221  CCTGTGAGAACAAGCAGCAGACCCGCTGTCCCGGGAGGCCGAGGTGAAGGCCCTGG      2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  CCTGTGAGAACAAGCAGCAGACCCGCTGTCCCGGGAGGCCGAGGTGAAGGCCCTGG      2280

Qy  2281  TCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACACCGCCAAGCAGC   2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  TCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACACCGCCAAGCAGC   2340

Qy  2341  TGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTACGGCGTCCCTCGCCGCCGCCGG      2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTACGGCGTCCCTCGCCGCCGCCGG      2400

Qy  2401  GTGTCCGGGAGGCCGGCGTCGTGTGGGCGGCGTACGGCGTACGGCGGTCCCGGAGATCCCG  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  GTGTCCGGGAGGCCGGCGTCGTGTGGGCGGCGTACGGCGTACGGCGGTCCCGGAGATCCACCG  2460

Qy  2461  GCGACGGCCACCCGTTCGGGTTACGCATCACCTACCTCCACCAGCCCGCCGCTCCGTC    2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GCGACGGCCACCCGTTCGGGTTACGCATCACCTACCTCCACCAGCCCGCCGCTCCGTC    2520

Qy  2521  TCGGCCACGCGGTGCCGCATCGCCGGCGGCTTCCTGGCGACGACGACTTCCTGCTGTACC  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  TCGGCCACGCGGTGCCGCATCGCCGGCGGCTTCCTGGCGACGACGACTTCCTGCTGTACC  2580

Qy  2581  TGGGGGACAACTACCTGCCCCAGGGCGTCGCCCGACTTCGCGGCGGCCAATCGGCCGCCGATC  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TGGGGGACAACTACCTGCCCCAGGGCGTCGCCCGACTTCGCGGCGGCCAATCGGCCGCCGATC  2640

Qy  2641  CCGGGGGGCCGCCCCGGCTGCTCACCCGGCTCACCCCGTCGCGG      2700
          |||||||||||||||||||||||||||||||||||||||||||||
Db  2641  CCGGGGGGCCGCCCCGGCTGCTCACCCGGCTCGCCGTCGCGG      2700

Qy  2701  AGGTCGACGCGGACGGGAACGTGCTCGCGCTTGGAGGAGAAACCCGGCGTCCGCGCAGCT  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  AGGTCGACGCGGACGGGAACGTGCTCGCGCTTGGAGGAGAAACCCGGCGTCCGCGCAGCT  2760

Qy  2761  CGCTCGCGCTCATCGGCGGTGTACGGCGTCAGCGCCGGCCGGCGGTACGGGCCA     2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  CGCTCGCGCTCATCGGCGGTGTACGGCGTCAGCGCCGGCCGGCGGTACGGGCCA     2820

Qy  2821  TCACCCCCTCCGCGCCCCGCGCCGGAGCTGGAGATCACCCGCGCGTCAGTGGATCGACC    2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  TCACCCCCTCCGCGCCCCGCGCCGGAGCTGGAGATCACCCGCGCGTCAGTGGATCGACC    2880

Qy  2881  GGGGGCTGCGGCGTACGGCGTGAGACCACCACCCGGCCCTGGCGCCGCAGCCGCGG    2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  GGGGGCTGCGGCGTACGGCGTGAGACCACCACCCGGCCCTGGCGCCGCAGCCGCGG    2940

Qy  2941  AGGACATGCTGGAGGTCAACCTGAGGTCCTGGACGACTCAGCGTCGTCACGCGCGAGGGA  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  AGGACATGCTGGAGGTCAACCTGAGGTCCTGGACGACTCAGCGTCGTCACGCGCGAGGGA  3000

Qy  3001  AGGTCGACGGGACCGGTACACCTCCATCGGGGAGGACTGCCCGGTGGCCGAAGGCCATCGTGC  3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  AGGTCGACGGGACCGGTACACCTCCATCGGGGAGGACTGCCCGGTGGCCGAAGGCCATCGTGC  3060

Qy  3061  GTGTCGACGCGCCCCGCTGCTCGCGGGGGCTCTGCTCTGCGGGGGCTCAGCAACTCCA    3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  GTGTCGACGCGCCCCGCTGCTCGCGGGGGCTCTGCTCTGCGGGGGCTCAGCAACTCCA    3120

Qy  3121  ACTCCGTCCGTCCTGCTCGCGGGGCTCGAGGCGGTCCGTCCATCGAGGTCCGTCCCTCA     3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121  ACTCCGTCCGTCCTGCTCGCGGGGCTCGAGGCGGTCCGTCCATCGAGGTCCGTCCCTCA     3180

Qy  3181  ACTCCGTCCGTCCTGCTCGCGGGGCTCGAGGCGGTCCCATCGAGGTCCGTCCCTCA     3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  ACTCCGTCCGTCCTGCTCGCGGGGCTCGAGGCGGTCCCATCGAGGTCCGTCCCTCA     3240

Qy  3241  TCGGCGGCGGCGTCGTCGGCCCCGCCGCGCGCCCGCCCGCTCCCGCCAGGCTCACCGACTGGTGA  3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  3241  TCGGCGGCGGCGTCGTCGGCCCCGCCGCGCGCCCGCCCGCTCCCGCCAGGCTCACCGACTGGTGA  3300

Qy  3301  TCGGCGGACCACAGCAAGGTGTATCTCACCCCATCGACCACGACACCTCCGTCGTGCCGGCCGG  3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  TCGGCGGACCACAGCAAGGTGTATCTCACCCCATCGACCACGACACCTCCGTCGTCACCGGGCGG  3360

Qy  3361  AGCGGGCTTCATTCGTGCCCGCTCGGCGTTCGCGGCGCTCCGCGGCCGGGGGCCCCGGGCGG  3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361  AGCGGGCTTCATTCGTGCCCGCTCGGCGTTCGCGGCGCTCCGCGGCCGGGGGCCCCGGGCGG  3420

Qy  3421  CGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCGGCAGCCTCGCCGCCTGCACGC     3480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421  CGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCGGCAGCCTCGCCGCCTGCACGC     3480

Qy  3481  GGTGCGTGACCATCCCGGCCTCACCTTGTCCCAGGGCGACGTGTGCACACCGGCGGCTCGT    3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  GGTGCGTGACCATCCCGGCCTCACCTTGTCCCAGGGCGACGTGTGCACACCGGCGGCTCGT    3540

Qy  3541  CGACACGCTGGCCGCGGCGACGACATCGTGCACTTCGCGGCGCTCGTCGCACGTCGA     3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  CGACACGCTGGCCGCGGCGACGACATCGTGCACTTCGCGGCGCTCGTCGCACGTCGA     3600

Qy  3601  CCGCTCCATCACCGACAGCGGTGCCTTCACCCGACCAACGTGCTGGCACCCAGGTCCT    3660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  CCGCTCCATCACCGACAGCGGTGCCTTCACCCGACCAACGTGCTGGCACCCAGGTCCT    3660

Qy  3661  GCTCGACGGCCGCGTCCGCGACTTCGTGCGCACCTTCGTCGTGCACGTCTCCACGCGACGAGGT  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  GCTCGACGGCCGCGTCCGCGACTTCGTGCGCACCTTCGTCGTGCACGTCTCCACGCGACGAGGT  3720

Qy  3721  GTACGGCGTCCCTCCCGCCGGCGGCGGACCCCGCTTCGGACCTTCGACCTCGCC     3780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  GTACGGCGTCCCTCCCGCCGGCGGCGGACCCCGCTTCGGACCTTCGACCTCGCC     3780

Qy  3781  GTACGGCGTCATGGCGCTCATGGCGGCCTCGAAGGCGCTCGAAGGCGTCCGCGCAGT   3840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781  GTACGGCGTCATGGCGCTCATGGCGGCCTCGAAGGCGCTCGAAGGCGTCCGCGCAGT   3840

Qy  3841  CCTGACGTCCGGGTGACCCGCTGTTCGAACAACTTCGGCCCCCACCAGATCCCGAGAA  3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841  CCTGACGTCCGGGTGACCCGCTGTTCGAACAACTTCGGCCCCCACCAGATCCCGAGAA  3900

Qy  3901  GCTCATACCGGCGCTTCCTGACCAGCCTCCTGTCGACCGCGTCCCTCTACGGCGA   3960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901  GCTCATACCGGCGCTTCCTGACCAGCCTCCTGTCGACCGCGTCCCTCTACGGCGA   3960

Qy  3961  CGGGGGGCACGTGCGACTGGCTCGACGACGTCGAGCACGTCGAACTCGT       4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961  CGGGGGGCACGTGCGACTGGCTCGACGACGTCGAGCACGTCGAACTCGT       4020

Qy  4021  CCGCGTGTCGGGGGCCGCGGGAGAGATCTACAACATCGGGGCGGCGGCCCGCCACCTGCTGCCCAA  4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021  CCGCGTGTCGGGGGCCGCGGGAGAGATCTACAACATCGGGGCGGCGGCCCGCCACCTGCTGCCCAA  4080

Qy  4081  CCTGGAGCGTTGCTTCTCGCACTGTGCCGCCGGTTCCCCCTCTACGGCGA       4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081  CCTGGAGCGTTGCTTCTCGCACTGTGCCGCCGGTTCCCCCTCTACGGCGA       4140

Qy  4141  CGTCGAGAACCGCAAGGGGCACGAGGACTGGTGCGCACGGCGCTACGGCGTCGAACTCGT    4200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4141  CGTCGAGAACCGCAAGGGGCACGAGGACTGGTGCGCACGGCGCTACGGCGTCGAACTCGT    4200

Qy  4201  GGAACTCGGTTACCGGCGGCCGCACTTCGGCGACCGCGGCGAAGTG      4260
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  4201  GGAACTCGGTTACCGGCGGCCGCACTTCGGCGACCGCGGCGAAGTG      4260

Qy  4261  GTACGAGGCGCACGAGGACTGGTGCGCCGTCGCGGACAATGACGTCGGGGCCGG     4320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4261  GTACGAGGCGCACGAGGACTGGTGCGCCGTCGCGGACAATGACGTCGGGGCCGG     4320

Qy  4321  ACCGAACCACCCGCGCGGCGGCCCGCCCACGCCGTCGCGGCGGCCGTGGGCCGGTCAG  4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Right column (top block):

```
Qy   5461 GCGGAACCGGCGGACTGCGGGCCACGCCCGTCCGCGGCACGACCACGCGCCCCGAAA 5520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 GCGGAACCGGCGGACTGCGGGCCACGCCCGTCCGCGGCACGACCACGCGCCCCGAAA 5520

Qy   5521 CGATGCTGAAAATGCTTGAGCAAATTGCCGCAACGCTCTTTCGGCGCTTTCGATCCT 5580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 CGATGCTGAAAATGCTTGAGCAAATTGCCGCAACGCTCTTTCGGCGCTTTCGATCCT 5580

Qy   5581 GTTACGTTCCTGGCAACCCGCGGCCGCCGCGGAGAAGCGGTTGGCGTGAGGCGTCCAGACCT 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 GTTACGTTCCTGGCAACCCGCGGCCGCCGCGGAGAAGCGGTTGGCGTGAGGCGTCCAGACCT 5640

Qy   5641 CCGCCCGATTCCGGATCACTCAGGGGAGTTCACAATGCGGCGTGGCATTGCGCAACGGCCGAA 5700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 CCGCCCGATTCCGGATCACTCAGGGGAGTTCACAATGCGGCGTGGCATTGCGCAACGGCCACCG 5700

Qy   5701 CGCTGTTCGCGGCTGTGGCCATGACGGCATCGGCGTGTGCGCACGGACCGTCACCTTCGTGGGGACCGTCACGTCCA 5760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 CGCTGTTCGCGGCTGTGGCCATGACGGCATCGGCGTGTGCGCACGGACCGTCACCTTCGTGGGGACCGTCACGTCCA 5760

Qy   5761 GCGGTACCGACGCGGCGGCGCACGGCGCACGTCGTCGGGGACCGTCACCTTCTGGGACGAGGAGCACCCGA 5820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 GCGGTACCGACGCGGCGGCGCACGGCGCACGTCGTCGGGGACCGTCACCTTCTGGGACGAGGAGCACCCGA 5820

Qy   5821 ACGAAGCCGAGAAGGCGACGTACCAGGCCGCCCTCGCGGAGGCTTCGAGAAGGAGCACCCGA 5880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 ACGAAGCCGAGAAGGCGACGTACCAGGCCGCCCTCGCGGAGGCTTCGAGAAGGAGCACCCGA 5880

Qy   5881 AGGTCGACGTCAAGTCGCAACTCCCGTTCGGCGAGGCGAACGCCAAGTTCAAGAACG 5940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5881 AGGTCGACGTCAAGTCGCAACTCCCGTGCCCGTCGGCGAGGCGAACGCCAAGTTCAAGAACG 5940

Qy   5941 CCGCGGGGGCCAACTCCGGTGCCCGTCGTCGTGCTACCTCCGGACGCACGCCCCGCGCGG 6000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 CCGCGGGGGCCAACTCCGGTGCCCGTCGTCGTGCTACCTCCGGACGCACGCCCCGCGCGGG 6000

Qy   6001 ACTTCGCGCAGCATCGGCTACCTGCGGGCCTCGGTCTCCGGTCGTCGGCGGACGGGT 6060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 ACTTCGCGCAGCATCGGCTACCTGCGGGCCTCGGTCTCCGGTCGTCGGCGGACGGGT 6060

Qy   6061 CGGACCACCTTCCCCAGGTACGACCGGCAGCCAGGTACGAGGGAGAAGGAAGACCTGACGAGGGTCGTCGGCGCGGTGTCG 6120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 CGGACCACCTTCCCCAGGTACGACCGGCAGCCAGGTACGAGGGAGAAGGAAGACCTGACGAGGGTCGTCGGCGCGGTGTCG 6120

Qy   6121 AGGTGATCGACACACCCTGGCGCTCTTCTACAACAAGGAACTGCTGACGAAGGCGGCGGACCG 6180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121 AGGTGATCGACACACCCTGGCGCTCTTCTACAACAAGGAACTGCTGACGAAGGCGGCGGACCG 6180

Qy   6181 AGGTGCCGGGCTCCTGCCGGAGTCACCGAGATCACCACCACCGACGACCG 6240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181 AGGTGCCGGGCTCCTGCCGGAGTCACCGAGATCACCACCACCGACGACCG 6240

Qy   6241 GGGCGAGCGGGCTCTCACCTGGGGGGGAAGAAGGGTTGTCCCGCCCACCTCTACG 6300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241 GGGCGAGCGGGCTCTCACCTGGGGGGGAAGAAGGGTTGTCCCGCCCACCTCTACG 6300

Qy   6301 GGGAGGGGCGGACCTGGTGCGACGAGAACAAGACCGTCACGTCGACGAAGCCG 6360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301 GGGAGGGGCGGACCTGGTGCGACGAGAACAAGACCGTCACGTCGACGAAGCCG 6360

Qy   6361 GTGTGCGCGCCTCACCCGCGTCATCGGCCATCACCGACG 6420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361 GTGTGCGCGCCTCACCCGCGTCATCGGCCATCACCGACG 6420

Qy   6421 CGTCCGACGGTGGAACAACATGCAGAGCTTCAAGGTCGGCGGCCAAGGTCGCCATGATGG 6480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421 CGTCCGACGGTGGAACAACATGCAGAGCTTCAAGGTCGGCGGCCAAGGTCGCCATGATGG 6480

Qy   6481 TCAACGGCCCCTGGGCGGCCATCGGAGGCGCCCGCCTTCAAGGACGGAGCGG 6540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 TCAACGGCCCCTGGGCGGCCATCGGAGGCGCCCGCCTTCAAGGACGGAGCGG 6540
```

Left column (bottom block):

```
Db   4321 ACCGCAACCACCGGCCGGCCACCGCCGCGGCCGTGCCGGCCGGTCAG 4380

Qy   4381 CGTCCGTGAGCGGACCGGCGGCCGGCCGCCCCGGGCCGGTGGACCACC 4440
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 CGTCCGTGAGCGGACCGGCGGCCGGCCGCCCCGGGCCGGTGGACCACC 4440

Qy   4441 AGTTCCGGCATGAAGACGAATTCGGTGCGGCGGCGCTTCCGCTCATCTCCTCCAGC 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 AGTTCCGGCATGAAGACGAATTCGGTGCGGCGGCGCTTCCGCTCATCTCCTCCAGC 4500

Qy   4501 AGTGGCGTCCACGGCGACCTGCCCCATGCGTCTGATGGTGGTCAGGGGA 4560
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 AGTGGCGTCCACGGCGACCTGCCCCATGCGTCTGATGGTGGTCAGGGGA 4560

Qy   4561 GGGTCGGTGAAGGCCATGAGCGGGGAGTCGGACCACCGAGATGTCACCGGGA 4620
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 GGGTCGGTGAAGGCCATGAGCGGGGAGTCGGACCACCGAGATGTCACCGGGA 4620

Qy   4621 ACCGTGAGACCCCGCGGGCCATCAGCGGCCATCATGTCGCTGGCG 4680
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 ACCGTGAGACCCCGCGGGCCATCAGCGGCCATCATGTCGCTGGCG 4680

Qy   4681 CACATGACGGCGGTGCAGCCCAGTCGATCAGCGGCCTCCCGACTCCCAGGTGC 4740
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 CACATGACGGCGGTGCAGCCCAGTCGATCAGCGGCCTCCCGACTCCCAGGTGC 4740

Qy   4741 AGGGAGAACAGCGAGTGCTGCACGAGCTCCTCCGGCACTCCCAGGTGC 4800
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 AGGGAGAACAGCGAGTGCTGCACGAGCTCCTCCGGCACTCCCAGGTGC 4800

Qy   4801 TCCCGCACGCCGCGGAACCCTCGTGCTGTGCCCGACTCTTCCGCTGCCCACCGGGCGGGGC 4860
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 TCCCGCACGCCGCGGAACCCTCGTGCTGTGCCCGACTCTTCCGCTGCCCACCGGGCGGGGC 4860

Qy   4861 CCGACGGGCGAGGCGGAGGCCGACGCGCACCGCGTGCGGAGAGCAGAAGGTGCCTCGATCTCGGGGAGCAGAAGGTGCCTCGGGGC 4920
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 CCGACGGGCGAGGCGGAGGCCGACGCGCACCGCGTGCGGAGAGCAGAAGGTGCCTCGATCTCGGGGAGCAGAAGGTGCCTCGGGGC 4920

Qy   4921 ATCGCGGCCGGTGTCGTCGGGGAGACGAAGAGGGTGCCTGATCCGGGGGCGGAACCCGTTC 4980
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 ATCGCGGCCGGTGTCGTCGGGGAGACGAAGAGGGTGCCTGATCCGGGGGCGGAACCCGTTC 4980

Qy   4981 ACGAGGACGAAGGGCACTGCGCCGTCGTGCGAGCCGGCCTGCGCCGTACCGTCCGGTCTCCGGCCGGTG 5040
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ACGAGGACGAAGGGCACTGCGCCGTCGTGCGAGCCGGCCTGCGCCGTACCGTCCGGTCTCCGGCCGGTG 5040

Qy   5041 GTGTCCGCGTGAGTCGCCGTGCGAGCGAGATGATGCCGGGGTCACACCCGGTCACGGAGCATC 5100
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 GTGTCCGCGTGAGTCGCCGTGCGAGCGAGATGATGCCGGGGTCACACCCGGTCACGGAGCATC 5100

Qy   5101 TCCGTGAGTTCGTCCTCGGCCTGGCCTTGCTCGTGTCGTGGCGCACGGGCCGTGTAG 5160
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 TCCGTGAGTTCGTCCTCGGCCTGGCCTTGCTCGTGTCGTGGCGCACGGGCCGTGTAG 5160

Qy   5161 CCCTGACGCGTGAGCCGCTGAGCGGCCTGGGGGAAGAAGGGTTGTCC 5220
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 CCCTGACGCGTGAGCCGCTGAGCGGCCTGGGGGAAGAAGGGTTGTCC 5220

Qy   5221 AGTTCGGGGGTGACCAGCTCCGACCAGTCGCGGCGTCGCGGCTGCTCGTAG 5280
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 AGTTCGGGGGTGACCAGCTCCGACCAGTCGCGGCGTCGCGGCTGCTCGTAG 5280

Qy   5281 CCCAGGCGCTGCGGTCGCGGAGTCGCCACCGGCGGTGCCGGCCACACCGCGCA 5340
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 CCCAGGCGCTGCGGTCGCGGAGTCGCCACCGGCGGTGCCGGCCACACCGCGCA 5340

Qy   5341 CGGTTCAGCACCGGCTGCACGGTCTGCTTGGCTTGGCCGTGCGATGTCGGGCG 5400
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 CGGTTCAGCACCGGCTGCACGGTCTGCTTGGCCGTGCGATGTCGGGCG 5400

Qy   5401 AGCCCGACGGTCATGCAACGCACTCTACCGTTCGGGGCGTGCGCCACCGGCGCGGCA 5460
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 AGCCCGACGGTCATGCAACAACATCTACCGTTCGGGGCGTGCGCCACCGGCA 5460
```

```
Qy   6541  ACCTGGGGTCGCCCCCGTCCCGGCCGGCAGCAGGGCTCTCCCAGGGCGGGT  6600
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541  ACCTGGGGTCGCCCCCGTCCCGGCCGGCAGCAGGGCTCTCCCAGGGCGGGT  6600

Qy   6601  GGAACCTTCGGTGTACCGGCGGCTCGAAGAACCTGGAAATAGAACTCGACTATGCGGTCGACGACG  6660
           ||||||||||||||||||||||||||||||||||| ||| |||
Db   6601  GGAACCTTCGGTGTACCGGCGGCTCGAAGAACCTCGACGCCTCTACGGTCGTGAAGT  6660

Qy   6661  ACATGAGCTCCGCCAAGGTGCAGCAGCAGCACCACCGAGAGCTCAGCCTCTGCCACCC  6720
           ||||||||||||||||||||||||||||||||||||||
Db   6661  ACATGAGCTCCGCCAAGGTGCAGCAGCAGCACCACCGAGAGCTCAGCCTGCTGCCACCC  6720

Qy   6721  GCACGTCCGTCTACGAGTCCCGTCGCGGACAACGAGATGGTGAAGTTCTTCAAGC  6780
           ||||||||||||||||||||||||||||||||||||||
Db   6721  GCACGTCCGTCTACGAGTCCCGTCGCGGACAACGAGATGGTGAAGTTCTTCAAGC  6780

Qy   6781  CGGCCGTGCACAAGGCCGTCGAACGGCCCTGGATCGCCGGAGGCGAACGTGCTGCCGTTGAGGAGAAAC  6840
           |||||||||||||||||||||||||||||||||||||||
Db   6781  CGGCCGTGCACAAGGCCGTCGAACGGCCCTGGATCGCCGGAGGCGAACGTGCTGCCGTTGAGGAGAAAC  6840

Qy   6841  CGATCCGGCTGCAG  6854
           ||||||||||||||
Db   6841  CGATCCGGCTGCAG  6854
```

RESULT 2
US-09-320-878-22
; Sequence 22, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: McDANIEL, Mary C.
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 3292
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-22

Query Match     7.7%;  Score 529;  DB 3;  Length 3292;
Best Local Similarity   53.4%;  Pred. No. 2.5e-68;
Matches 1476;  Conservative    0;  Mismatches 1180;  Indels 109;  Gaps    13;

```
Qy   2269  TGAAGGCCCTGGTCCTGGCAGGTGGAACCGGCAGCAGCCGTCACCCACCACCG  2328
           |||||||||||||||| || ||| ||||||| |  |||
Db    105  TGAAGGGAATAGTGTCTCGGCGGCGGGCGGCGAACTCGGCATCGGCCACCTCGGTCA   164

Qy   2329  CGGCCAAGCAGCTGCTCTCCCCATCGCCAACAAGCCCGTGCTTCTTACGCGTGGCGTGAGTCCC  2388
           ||||| |||| || |  ||||| ||| ||  |
Db    165  TTTCGAAGCAGATTCTTCCGGTTCCGGAGGCCCCG------GCCGGTGCGGGCCGCCTCGGAGTCGC   224

Qy   2389  TCGCCGCCGCTGAGGTGTCCCGGGCGGTGCGGGAGGCCGCGTGCCGTACGGCCGCGGAGATCC  2448
           ||  |||| ||| ||  || ||||  ||  ||
Db   1236  CGCGGGCAACCCGGCCAACCTCGCCCGGACTCGGGCGACTCGCCGGGACCGGCCTTCGTCCA  1295
```

```
Db    225  TCATGCTCGCGGGTATTCGGGAGATTCAAATCATCTCGACCCCCAGCACATCGAACTCT   284
Qy   2449  GCGA---ACTCACCGCGACGGCACCGTTCGGGTTACGGCTACCTCCACCAGC  2505
           ||| |  ||| ||
Db    285  TCCAGTCGCTTCTCGAAACGCAGCAGCACCTGGAATAGAACTCGACTATGCGGTCGACGA   344
Qy   2506  CCCGCCCGCTCGGTCTCGCGGCGGTGCGCGACTTCTCGTCGGAGCGCGAGCAGCATCGGCGACGACCC  2565
           ||||| ||| ||  ||| ||| ||||  ||
Db    345  AAGAGCCCAGGAAGAATGCGGACGCGACTTCTCGTCGGAGCGCGAGCAGCAACTACCTGCCC   404
Qy   2566  ACTTCCTGCTTACCTGGGGGACAACTACCTGCCC---CAGGGCGTCACCGACTTCGCCC  2622
           ||| |  |||||| ||| ||  ||| |||| ||
Db    405  CCTGCGCCCTGATCCTGGGCGACAACATCTTCCACGGGCCGGCCGGTCACCCCGGTCGCGGACCGC   464
Qy   2623  GCCAATCGGCCGCCGATCCCGCGGACGCGGCGGCCGCTCGCTCACCCCGGTCGCGGACCGC  2682
           ||| |||  || ||| || || ||||  ||
Db    465  GGGACAGCATCGCCGGCCCTGGATCGCCGGAGTCGGACGCGGAACGTCGTCGGCGACGGCAATGTCGCC  524
Qy   2683  CGCCTTCGGCCGGTCGCGGAGGTCGACGGCGACGCGGCTGCGGGCCGGAACGTGCTGCCGTTGAGGAGAAAC  2742
           |||  |||||||  || ||||||| |  |||
Db    525  AGCGGTACGGCGTCGCGGGTGACGCGGCGGCTGCGACCCGGCTCGTGACGCCGTCC   584
Qy   2743  CCGAGCTCCCCGCAGCTGCTCGCTCGGCGCTCATCGCCGCTCAGGCTGTAGCGGTCGTCGCCTCC  2802
           |||  ||| |||  | |||
Db    585  CCGTCAAGCCGCGCTCCAACCTCGCGCCTCACCGCTGGCCGAGCTCATCTACGACAAACGAGGTCG   644
Qy   2803  ACGAGGGGTACGGCGCTCCGGGAGTCCGACCGGAGCTGGAGATCACCAGCCCG  2862
           ||||||||||| |  ||||| |  |||||
Db    645  TCGACATCGGCCAAGAACATCCGCCCTCGCCGCGGAGATCACCGACGTCA   704
Qy   2863  TGCAGTGGATGATCGACACCGGGGCCTGCGGGTACGGGCCGA--GACCACCACCCGGCCCT  2919
           |||| ||| || ||  || ||||  ||  ||
Db    705  ACCGGGTCTACCTGGACGGGCCGGCGGAACTCGTCAACCTGGGCCCGCGCTTCGCCT   764
Qy   2920  GGGGCGACACCACCGGCAGCGCGGAGGACATGCTGGAGTCAACCGTCAACGTCCGTGGACGGAC  2979
           ||  ||||| ||  ||||||| |||| | |||
Db    765  GGCTGGACACCGCGCACCCACGGTGGCAGCCGCCCAGTACGTCCAGGTCCTGG   824
Qy   2980  TGGAGGCCGCATCGAGGGAGAGGTCGACGCGGCCCACAGCACGGTGGTCGGCCGGGTCCGGG  3039
           ||| |  ||| ||  || ||||
Db    825  AGGAGCG--------------GCAGGGCGTCTCGGATCGCGGCGGCGGGGTCTGG   851
Qy   3040  TGGCCGAAGGCGCGATCGTGGCGGGGGTCACACGTGGTGGGCGCCGGCCTGTCGGGGCGTGG  3099
           ||| || ||| ||  || ||||  ||
Db    852  GCCTTGAGGAGATCGCCTTCCGCCATGGGCTTCATCGACGCCGGCCTTGCGCGGCGGGCCTGG   911
Qy   3100  GTGCCGTCGTCAGCAACTCCAGTGTCCGCCTACACCTTCATCGGGGAGGACGGACTGCCGGGG  3159
           || | |  || |  ||   ||
Db    912  GAGAAGCCCTCTCCCCGCCCAGTAGCGCAGCTA-------TCTGATGGAGATCGCGCGGC   964
Qy   3160  TCGAGGACAGCGCCATGAGCTACTCCTCGTCCTCGTCGCCGGCCCCAGTGCGAGGGGGCGT  3219
           |  || ||||| | || |  | |   |||
Db    965  CGGCGAGGGGAGCCCCGTGAGGCGACCACGCTCGGCGGCCGCCGACGCGTTCCCGACCAGCCCA  1024
Qy   3220  CCCGCATCGAGGCGCTCCCTCATCGGCCGGACCACCGCGGCCGCACCGGCCACCGACAGTGCGACC  3279
           |||  ||| || |||||||  |  ||| ||| || ||||||
Db   1025  CCGACAGTGCGACCACCGCCGCCGGCCACCGCCACCGGCGACC---------------  1074
Qy   3280  CGGCAGCCTCACCGACTGGTCAGGCTCGCGGCGACCACAGCAAGGTGTATCGCCGCCATGACCAC  3339
           |||  ||||   ||  |  ||
Db   1075  ------CACACCCGGACCTACAGCGCGGACGCGGACGCGAAAGGAAG-----------ACGGCAGT  1115
Qy   3340  GACCATCCTCGTCACCGGACCGGCGAGCGGAGGCGGCTTCATTCGCTCGCAGCTCCGGCTCCT  3399
           ||||| |||| || |||  || ||| || || |  ||
Db   1116  GCGGCTTCGGTGACCGGAGGTGCGGCGTGCGCCACCTTCGCGGCTCCT  1175
Qy   3400  GTTCGCCGTCGCGGCGCCCCG------GCGGCGTGCGGCCCCGTCCCGGTGACCGCTGGAGTCCC  3453
           ||| || ||  ||||| |||    |   | |||  || |||  |||
Db   1176  CGCCGGGGGCTACCCCGCGTGCCCGGATGAGGTGATCGTCCTGGACGGCCTCACCTA  1235
Qy   3454  CGCCCGCCTCGCCCTGCCTCCCCGGCCTCGCACCGCGGCCTCACCTTCGTCCA  3513
           |||||| ||||  ||  || ||| |||  ||
Db   1236  CGCGGGCAACCCGGCCAACCTCGCGCCTCACCGCTGGCCGAGCTCATCTACGACAAACGAGGTCG   1295
```

```
Qy  4573  GCCATGAGCGGCGAGTCGTCGAAGCCGACCACCGAGATCTCACCGGGAACCGTGAGACCC  4632
          |||   ||   |     |||   | |      | |  |    | |  ||   |||
Db  2376  GAGGTGATCGTCCCCTCGCACACGTACATCGCCAGCTGGCGTCGCGGTGTCGCCACCGGC  2435

Qy  4633  CGGCGGCCGGCCGGCCGCCGGGCGCGAGGGCCATCATGTCGGCGGCACATGACGGGCG  4692
          |  |  || |  |||  || |   ||  ||
Db  2436  GCGACCCCGTGCCGTCGAGCCGCCGACGAGGACCACCCCACCCTGCCCCCGTCCCTCGTC  2495

Qy  4693  GTGCAGCCCAGGTCGATCAGCGCGGACGCGG-CGGCCTGCCCCCTCCAGGGAGAAACAG  4751
          |  |  | |||  |      |   | | |      |||   |  ||| | |  |
Db  2496  GAGAAGGGGATCACCACCCCCGGGCGCTCCTCTACGGCCTCTACGGCCACCCC  2555

Qy  4752  CGAGTGCTGCACGAGCTCCTCGGCGACTCCCGCCGCCGACTCCCAGGTGCTCCCGCACGCC  4811
          |  |    ||  ||   |   || ||  || |  ||  |  | |  | |
Db  2556  GCCGACATGACGCCCTCGATCTTCCGCGAGCCGCTCGGCGTCGGCGACCGGGGGGGGAG  2615

Qy  4812  GGCCCGGAACCCCTGATCTTCCGCTCCCGCCAGTGCCGCCACGCGGCCGCGGCCATCGGGCCCG  4871
          | |     |     | |   || |  |  ||  ||     ||  |||
Db  2616  GCCCCGCAGGCCACG--GCGCCCGCTACCGGGCGCCGGCGCGGCTGCTTCGGGGGGCTCGTCG  2672

Qy  4872  GCCGACGCGCTCGTCGCCCAGTCCCGCCACGTCGCCCAGTGCGCCAGCGGCCATCGGCGGCCCG  4931
          |  |  |   |      |      ||   |  |  | | | | |  |||| |
Db  2673  GTGGCCGCGTTCAGCTTCACCGGCCCAAGAACCTCGGCTGCTTCGGGGGCAGCGGGGGGC  2732

Qy  4932  GTCGTCCGGGGAGACGAAGGGTGCCTCGATCCGCGGGCGAGAACCCGTTCACGAGGACGAA  4991
          |  |  | |  |  |  | |   |||
Db  2733  GTCGTCACCGGGCGGCACCCCGAGCGTCGCCGACGGCTCGCGATGCTCCGGAACTACGGCTCG  2792

Qy  4992  GGGCA  4996
          |||||
Db  2793  CGGCA  2797

RESULT 3
US-09-105-537-3/c
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;   APPLICANT: Sherman, D.H.
;   APPLICANT: Liu, H.
;   APPLICANT: Xue, Y.
;   APPLICANT: Zhao, L.
;   TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;   FILE REFERENCE: 600.438US1
;   CURRENT APPLICATION NUMBER: US/09/105,537A
;   CURRENT FILING DATE: 1998-06-26
;   NUMBER OF SEQ ID NOS: 43
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 13613
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-3      Length 13613;

Query Match            7.7%;   Score 527.4;  DB 4;   Length 13613;
Best Local Similarity 53.3%;   Pred. No. 4.2e-68;
Matches 1475;  Conservative    0;  Mismatches 1181;  Indels 109;  Gaps  13;

Qy  2269  TGAAGGCCCTGGTCCTGGCAGGTGAACCGGCAGCGACTGAGGCGGTTCACCCACACCG  2328
          ||||||  |  || ||  ||||| |   || | ||  ||| |||| |     |
Db  9988  TGAAGGGAATAGTCCTGGCCGGCGGAGCCGGAACTCGGCTGCATCCGGGCGACCTCGGTCA  9929

Qy  2329  CGGCCAAGCAGCTGCTCCCATGCGGCCGACCACGGCGCGTCTTCTACGGCGTCGGAGTCCC  2388
          |  | |  | |  ||  |     ||||  |  |  ||  |||  || || |||||  |
Db  9928  TTTCGAAGCAGATTCTTCGGTCTCACAACAAACCGATGATCTACATCCGCTGTCGGGGTC  9869

Qy  2389  TCGGCCGGCGGGGTCGGCGGTGTCCCGGAGGCCGTCGTCGTGGGCGGGCTACGGCCGGAGATCC  2448
          ||||| |  ||   |  | || |  ||||| ||  | | |  || |||| |
Db  9868  TCATGCTTCGGCGATTCGGCACCGCGTTCGGGTTACGCACCTCGGGATTCAAATCATCTCGAACATCGAACTCT  9809

Qy  2449  GCGA--ACTCACCGGCGACGGACTCCTCGGGCCATCACCTACCTCCACCAGAC  2505
          | | |     |||  |  ||  || | || ||   |  |     |  | || |
Db  9808  AACAGCGGGATGGCGATGACGGACGTCCGCCCTCGCATCGGGAAATAGAACTCGGACTATGCGGTCCAGA  9749
```

```
Qy   2506 CCCGCCCGTCGGTCTGGGCTGCGCATCCCGCGGCGTTCCTGGGCGACGACG 2565        Qy   3574 GCACTTCGGCGGCCGAGTCGCACGTCGACCGCTCCATCACCGACAGCGGTGCCTTCACCCG 3633
Db   9748 AAGAGCCCGCAGGAATGCGGGACCACTTCTCGTCGGACGCACATCGGCACGACGACA 9689        Db   8737 CCACTTCGGCGGCCGAGAGCCACGTGGACGGCGTCCGCGCTCCGACGCGTCCTGTTCACCGA 8678

Qy   2566 ACTTCCTGCTGTACCTGGGGGGACAACTACCTGCCC---CAGGGCGTCACCGACTTCGCCC 2622      Qy   3634 CACCAACGTGCTGGGCACCCCAGGTCCTGCTGGGACGCGGTTCGGCGACGCGTCCTCGGCCAC 3693
Db   9688 CCTGCGCCCTGATCCTGGGCGACAACATCTTCCACGGCGTCCTACACGGTCCTGC 9629        Db   8677 GACCAACGTGCAGGGCACGCAGGCGCTGTCCAGTGGCCGTCCAGTGCGCGGGCGGGTCGGCCG 8618

Qy   2623 GCCAATCGGCCGCGCGATCCCGCGGCGGCCTGTCGTGCTGCACCCGGTCGCGACCCGT 2682      Qy   3694 CTTCGTGCGACGTCTCCACCGACGAGGTGTACGGCTCCTCCCCGCCCGCGGCGGA 3753
Db   9628 GGGACAGCAGCATCGGCGCGCCTGACGCGTGCGTCGTCGCGTCGTCTTTCGGCGTCAAGGACCCCCG 9569   Db   8617 GGTCGTGCACGTCCTCCACCGACGAGGTGTACGGCGGCGCTCCGGACTCCGGACCGA 8558

Qy   2683 CCGGCCTTCGGCGCTGCGGGAGGTCGACGCGGCAGCGGAGTGCTGCTGGAGGGGAAAC 2742     Qy   3754 GAGGCGACCCCTGCTTCCGACCTCGCCGCGTCAGGCGTCGAAGGCGGCCTCGGACTCAT 3813
Db   9568 AGCGGTACGGCGTCCCGAGGTGACGCGACGACGGCCGCTGACCGACCTCGTCGAGAAGC 9509    Db   8557 GAGCAGCCGCGTTGGAGCCCCAACTCGCCGGGCTGCCAAGGCCGCCGACCTCGT 8498

Qy   2743 CCGACGTCCCGGCAGCTCTCCGATCGGCGTGTACGGCTTCAGCCCGGCCGTCC 2802         Qy   3814 GGCGCTCGCCCACCACCACGCACCGCGCTCGGGGTGACCCGCTCGTTGTTCGAACAA 3873
Db   9508 CCGTCAAGGCCGCGGCTTCCCAACCTCGCCGTCTAACCTCTACGACACGACGTCG 9449        Db   8497 TGCCCGCGGCCTACCACCGACGTACGGCGCTCAGCGCGACGTGGGTCGCACCGA 8438

Qy   2803 ACGAGGCGGTACGGGCCATCACCCCTCGCGCCCTCGCGGCAGCTGGAGATCCACCCACGCCG 2862   Qy   3874 CTTCGGCGCCCCACCAGCATCCCGAGAAGCTCATACCGCGCTTCGACCAGCCTCCTG 3933
Db   9448 TCGGACATCGCCAAGAACATCGGCCGGCTCCACGTCCGCGCTGGAGATCACCGACGTCA 9389     Db   8437 CTACGGCGCCGTACCAGCACCGAGAGCTCATCCCCCTCTCGTGACGAACCTCCTCGA 8378

Qy   2863 TGCAGTGGATGATGACCGGCGGCCGCTGCGGCTACGGCCGGA---GACCACCACCCGGCCCT 2919   Qy   3934 CGGCGGCGCACCGTTCCCTCTACGGCGACGGCGCACGTGCCGCGACGTGGCTGCACGTCGA 3993
Db   9388 ACCGCGTCTACCTGGAGCGGCGGCAGGCCGAACTCGTCAACCTGGCCGCGCGGGCTTCGCT 9329     Db   8377 CGGCGGGACGCGTCCCGGCGTACGCGGCGAGTCGGGCGCGGCGGTCGCACCAGCA 8318

Qy   2920 GGGCGGACACCGGCAGCGGCCGCGAGAGACATGTCGGAGGTCAACGTCCTGGACGGAC 2979     Qy   3994 CGACCAGCGTCAGGGCCGGCGTGCGCGAACTCGGTCCGGGCGGCCGGGAGAGATCTACAA 4053
Db   9328 GGCTGGACACCGGCACCACGACTCGCTCCTGCGGGCCGCCAGTACGGTCCAGGTCCTGG 9269       Db   8317 CGACCACTGCCGCGGCATCGCCGGTCCTCGTCGCCGGGCGGCCGCCGAGATCTACCA 8258

Qy   2980 TGGAGGGCCGCATCGAGGGGAAGGTCGACGCGCGCACGCTCGTCGGCGTCGGGGTCGGGG 3039   Qy   4054 CATCGGGGCGGCACCTCGCTGCCAACCTCGGAGCTGTTGCTCTCGCCACTGTG 4113
Db   9268 AGGACCG----------------------------GCAGGCGGTCTGGATCGCGG 9242       Db   8257 CATCGGGCGGCGGCCTGGAGACCTGCCAACTCACCACCGGCTGCGACTCGCT 8198

Qy   3040 TGGCCGAAGGCGATCGTCGGGGTCACCGTGTCGGGGTGGTGTATCGGCCGGCGGCGGG 3099    Qy   4114 CGGCGCGGCCGCCGGAGCGCATCGTCCACGTCGCAGAACGTCGAGAGAACCGCCAAGGGCGCTA 4173
Db   9241 GCCTTGAGGAGATCGCCTTCGCATGGGCTTCATCGACAGCGCCGGAGCCGTTCACGGCCTG 9182   Db   8197 CGGCGCCGACTGTCCTCGGACGAAGGTCGCCCACGCACGTGCGCCCAAGGGCCGCGCTA 8138

Qy   3100 GTGCCGTCGTCAGCAACTCCAGTGTCGGCCCGTACACCTCCATCGGGGAGGACTGCCGGG 3159   Qy   4174 CGCGGGTGACCAAGAGATCACCGGCGGCGGAACTCGGTTACCGGCGGCCGCCACCGACTTCGC 4233
Db   9181 GAGAAGGCCTCTCCCCGACCGCCCGCCCACCGAGTCGGCAGCTA------TCTGATGGAGATCGCCGAC 9129   Db   8137 CTCCCTGACGCGACGGCGAGATCGAGCGCCGACGTCGGCTACCGCCCGAGAGTCTCCTTCGC 8078

Qy   3160 TCGAGGACAGCGCCATCGAGTACTCCGTCGTGCGCCCCAGGTCGAGGGGGCGT 3219         Qy   4234 GACCGCGCTGGCCGACACCGGCGAAGTGGTACGGCGGCGACGAGGACTGTCGTGGGCGTCCCCT 4293
Db   9128 CGGGAGGGAGCCCGTGAGGGCCGACCCGACGCGCCACAGCGCCA 9069                   Db   8077 GGACGGCCTCGGACGCCTCGACCGTGTACGCCGTGTGGGGAGCCCGCT 8018

Qy   3220 CCCGGCCATCGAGGGCGTCCCTCATCGGCCGCGCCGGCCGTCGTCGCGCCCCCGCGTCTCC 3279   Qy   4294 GCTCGCGCCGGCGACATGACGTCGG--------GCCGGACCCGCAACCACCGGCGACCGGCGCTA 4344
Db   9068 CCGACAGTCGACGTGCGACCCACACCGCGACCGCCACCGACAGTGCGACC--------- 9019   Db   8017 CAAGGCGGACGCCGCCCCGCTGCCGCCGTGGAGGTGTCGCGGCAGCCGCCGCTA 7958

Qy   3280 CGCAGGCTGTCGACGGTGATCGGCGACCAAGCGCCGACCTCACCCCATGACCAC 3339        Qy   4345 CACACCGCCCCCGCCGCGCCGCCGCGTTCCGTCGCCATCCTCTCCCGCTCAGGTGCCGGCCG 4404
Db   9018 --------CACACCGCGACCTACAGCGCCGACCGGAAAGGAAG 8978                    Db   7957 GCCGAGACCCCGTCCCCTTCCTGACCCGCCAAGGTGCTCCGGGGGCCGGCGCG 7898

Qy   3340 GACCATCCTCGTCACCGGGCGAGCGGCTTCATTCGCTCCGCCTACGTCGCCGGCTCCT 3399    Qy   4405 --------CAGTTCGGCGACATGAAGAACGAA 4460
Db   8977 GCGGGCTTCTCGTGACCGGCGGCGGGCGGCGGTTCATCGGCTGCCGCTCGCCGGCTCCT 8918   Db   7897 GAGACCGAGATCCGGGCGCCGGATCGGCCTCGGGGCGGCCGACCCGAA 7838

Qy   3400 GTCGCCCGGGCGCCCCG-----GCGGGTCGCGGCTCGCACGCGCTGCGTCGACCCGTG 3453    Qy   4461 TTCGGTGCGGCGCGGCGGGGCGTTCCTCATCGTCGCCGTCATCTCCTCAGCAGTGCGTCATCGCACCTG 4520
Db   8917 CTGGAAGGATTCGAGGCGGCGGCTGGCCTCCGGCCTCAACAAACTCACCTA 8858           Db   7837 CTGAAGGCTCGAGGCGGAGTTCGCGGCTACTGCGAGACGCGTACGCCGTGCGTGTG 7778

Qy   3454 CGGCGGCCGAGCGCTCGGCGCCCTGCAGCCATCCGCGGCCTCACCTTCGTCCA 3513         Qy   4521 CCCCCATCGCCTTGACGGGCT-------GTCTGATGGTGGTCAGGGGAGGGGTCGGTGAAG 4572
Db   8857 CGGCGGCAACCTCGCCTCCACCGCCGAACCTGGGCGACCTGCGCTTCGTTCCTCA 8798       Db   7777 AACAGCGGATGGACGACCCGCGCATCGGGCATCCGGGAACCGTGACCCGGGAC 7718

Qy   3514 GGGCCACGTGTGCGACACCGTCGTCGACGACAACGCTGGCCGCGCGCGACGACGACATCGT 3573   Qy   4573 GCCATGAGCGGCGGCGAGTCGTCGAAGCCGACGCACCGAAGATGTCACCGGAACCGTGAGACCC 4632
Db   8797 CGGCGGCGACATCGCGACGCGACGGCGCCGCGCGCGGAGCTCCGCGGGAGCTGCGGCGGCCATCGT 8738   Db   7717 GAGGTGATCGTCCCCTTCCGCGTCTCCGCGGCTGCGTCGTCCGCCACCCGGC 7658

                                                                        Qy   4633 CGCCGGCGGCGCGCGCGCCCCACGGCGGCCCATCATGTCCGCTGGGCGACATGACGCGCG 4692
```

```
Db  7657 GCGACCCCCGTGCCCGTCGAGCCGCCGCACGAGGACCACCCCCACCCTGGACCCGCTGCTCGTC 7598
              |  || |||  ||||    ||         ||    ||||| | ||||    ||   |

Qy  4693 GTGCAGCCAGGTCGATCAGCGGACGCGG-CGGCCTGGCCCCTCCAGGGAGGAACAG 4751
         ||| | ||  | | || |  ||      | |||    |  | || ||||  ||   |
Db  7597 GAGAAGGGGATCACCCCCGCACCCCGGCGCGCTTCCTCCCCGTCCACCTTCTACGGGCCACCCC 7538

Qy  4752 CGAGTGCTGCACGAGCTCCTCGGACTCCCGCCGCGACACTCCCAGTGTCCTCCCGCACGCC 4811
         |  | ||| ||||  |  || | || ||| |||   |||| |||  | || |  |
Db  7537 GCCGACATGGACGCCCTTCGCGACCGGCACCGGCCTGCACATCGTCGAGGAC 7478

Qy  4812 GGGCCGGAACCCCTCGATCTTCCGGTCCACCGGGCACCGGAGAAGGCGGGCGCCGACGGCGAG 4871
         ||| | |  || ||| || ||  |  | |||| ||    ||||| |||   |  ||| | |
Db  7477 GCCGCGGCAGGCCCACG---GCGCCCGTACCGGGGCCGGATCGGCCGCCGGGTCGTCG 7421

Qy  4872 GCCGACACGCGGCTCGTGCCCCCAGCTCGGCCCAGTCGCCCACGGCGCCAGGCGGCCGGCCCG 4931
         ||| | ||| || |  |   ||| || |   | ||  | ||  | ||||  |||| |  ||
Db  7420 GTGGCCGCGGTTCAGCTTCTACCCGGGCCAAGAACCTCGGCCTGCTTCGGCCGACGGCGGCGCC 7361

Qy  4932 GTCGTCCGGGGAGACGAAGGGGTGCCTCGATCCGGGCGAGAACCCGTTCACGGACGAA 4991
         ||||| |  |||  ||  ||  | |  || || ||   |||  | ||| ||   | |
Db  7360 GTCGTCACCCGGCGGACCCCCGAGCTCGCCGAACGGGCTCCGGATGCTCCGAACGGCTCG 7301

Qy  4992 GGGCA 4996
         |||||
Db  7300 CGGCA 7296
```

```
;  MOLECULE TYPE:  DNA (genomic)
US-09-194-905-3

Query Match          7.6%;  Score 522.6;  DB 4;  Length 546;
Best Local Similarity   97.4%;  Pred. No. 2.1e-67;
Matches 531;  Conservative   0;  Mismatches 14;  Indels   0;  Gaps   0;

Qy  3352 CACCGGCGGAGCGGGCTTCATTCGCTCCGCCTACGTCCGCCGTCCTGTCCGCCGGGGGC 3411
         |  ||||| |||| ||||||| ||||||||||||||||||||||||||||| ||||||||
Db     1 CCCGGCGGGGGCGGGGTTCATCCGCTCCGCCTACGTCCGCCGTCCTGTCGCCGGGGGC 60

Qy  3412 CCCGGCGGGCGTCGCGTGACCGTCCTCGACAAACTCACCTACGGCCGCAGCCTCCCCG 3471
         ||||||||| ||||||||||||||||||||||||||||||||||||||||||| |||||
Db    61 CCCGGCGGGGCGTCGCGTGACCGTCCTCGACAAACTCACCTACGGCCGCAGCCCTCCCCG 120

Qy  3472 CCTGCACGCGGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACAC 3531
         ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCTGCACGCGGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACAC 180

Qy  3532 CGCGTCTGTCGACACGCGTGGCCGCGGCGCACGACGACATCGTGCACTTCGGCGGCCGAGTC 3591
         |||||| ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   181 CGCGTCTGTCGACACGCGTGGCCGCGGCGCACGACGACATCGTGCACTTCGGCGGCCGAGTC 240

Qy  3592 GCACGTCGACCGTCCATCACCGACAGCGGTGCCTTCACCCGCACCGGCCAC 3651
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCACGTCGACCGTCCATCACCGACAGCGGTGCCTTCACCCGCACCGGCCAC 300

Qy  3652 CCAGGTCCTGCTCGACGCCGCGCTCCCGCCACGGTGCGCACCTTCGTGCACGTCTCCAC 3711
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCAGGTCCTGCTCGACGCCGCGCTCCCGCCACGGTGCGCACCTTCGTGCACGTCTCCAC 360

Qy  3712 CGACGAGGTGTACGGCGTCCCTCCCCCGCCGCGCGGAGAGCCGACCCCCTGCTTCC 3771
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CGACGAGGTGTACGGCGTCCCTCCCCCGCCGCGCGGAGAGCCGACCCCCTGCTCCC 420

Qy  3772 GACCTCGCCGTACGGCGCCTCGGAAGGGCGCCTCATGGCGCTCGGCACCACCACCG 3831
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACCTCGCCGTACGGCGCCTCGGAAGGGCGCCTCATGGCGCTCGGCACCACCACCG 480

Qy  3832 GACCCACGGCCTGGACGTCCGGGTGACGTGACCGGTGTTCGAACAACTTCGGCCCCCACCAGCA 3891
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   481 CACCCACGGCCTGGACGTCCGGGTGACGTGACCGGTGTTCGAACAACTACGGCCCCCACCAGTT 540

Qy  3892 TCCCG 3896
         ||||
Db   541 CCCGG 545
```

RESULT 4
US-09-194-905-3
; Sequence 3, Application US/09194905
; Patent No. 6306627
; GENERAL INFORMATION:
; APPLICANT: DECKER, Heinrich
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
; TITLE OF INVENTION: GLA.O AND THEIR USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194,905
; FILING DATE: 29-JUL-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP97/02826
; FILING DATE: 30-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19622783.6
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granados, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 026083/0193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

RESULT 5
US-08-196-218-31
; Sequence 31, Application US/08196218
; Patent No. 5614619
; GENERAL INFORMATION:
; APPLICANT: Piepersberg, Wolfgang
; APPLICANT: Stockmann, Michael
; APPLICANT: Taleghani, Kampiz Mansouri
; APPLICANT: Distler, Jurgen
; APPLICANT: Grabley, Susanne
; APPLICANT: Sichel, Petra
; APPLICANT: Brau, Barbara
; TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes
; TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their
; TITLE OF INVENTION: Use.
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20005-3315
; COMPUTER READABLE FORM:

```
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/196,218
;     FILING DATE: 25-AUG-1994
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Ogden, Stasia L.
;     REGISTRATION NUMBER: 36,228
;     REFERENCE/DOCKET NUMBER: 02481.1372-00000
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-408-4000
;     TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 31:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2634 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 3..401
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 416..1531
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1561..2625
US-08-196-218-31

Query Match        7.3%;   Score 497.6;  DB 1;  Length 2634;
Best Local Similarity  66.3%;   Pred. No. 8e-64;
Matches  716;   Conservative   0;   Mismatches 364;   Indels   0;   Gaps   0;

Qy   2256 GCGGAGGGCGGAGGTGAAGGCCCTGGTCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCG 2315
          |  |||| | ||||||||| |||  || ||||| ||||||  || ||| | |||| |||||
Db   1549 GAGGACCTTCGATGAAGGCTCTGTGCTCGCCGGATCTGGTACCCGCCCGTGCTCTTCTAC 1608

Qy   2316 TTCACCCACACGCGCCCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTTCTCTAC 2375
          ||||    || ||  |||  |||||| || ||||||  ||||||||||||||||| | |
Db   1609 TTCAGTTATTCGATGCCCAAACAACTGATCCCCATCGCCAACACACCCGTGCTGTGCAT 1668

Qy   2376 GCGCTGGAGTCCCTCGCCGCGGCGGTGTCGGGAGGCCGGCGTCGTGGGCGGCGTAC 2435
          | |||| | ||||||||  ||| ||||||||||||||||||| |||||| |||||
Db   1669 GTGCTGAACCCGTCGGGAGCTGGGCGTCGGCGTCATCGTCGGCGGCAACCGC 1728

Qy   2436 GGCGGGAGATCCGCGAACTCACCGGCGACGACGCACCGCGTTCGGGTTACCATCACCTAC 2495
          |||| ||||| || ||  |||||| || || ||  |  | ||||||||| |||||||||
Db   1729 GGCCCCGAGATCGAGGCCGTGCTCGGCGACGGTGCCCGGTTCGACGTGCGCATCACCTAC 1788

Qy   2496 CTCCACCAGCCCGCGCCGCCTCGGTCTCGCCGTCCCGACGGTGCGCCATCGCCCGGCTTCCTG 2555
          | ||| |||| |||||||||||  |||||||||  ||| || || ||||||||||||||||
Db   1789 ATCCCCCAGGACGACGCCACCGCGCGGACGGCCCACACCCGTGCCCATCGTGCCCGGCTTCCTC 1848

Qy   2556 GGCGACGAGACTTCCTGCTGTACCTGGGGACAACTACCTGCCCCAGGGCGTCACCGAC 2615
          ||||| |||| ||| | ||  |||| || | ||  | ||| |||  |||| |||| |
Db   1849 GGCGACGACGACTTCGATGTACCTCGGCGACAACATGCTCCGCGACGGAGTCACCCGAG 1908

Qy   2616 TTCGCCCGCCAATCGGCCCGCGCGATCCGCGGATCCGCCGCCGTCCTCCACCCCGGTCGCG 2675
          ||||  |  || || ||  | | | | |  | ||| | | ||||| ||  || | |||
Db   1909 ATCGCCGAGGAGTTCACCCGGCAGCGCCCGCCCGCCGAGTCGTCGTGCACAAGGTCCCC 1968

Qy   2676 GACCCCGTCCGCCTTCGCCGTCGCGGAGGTCGACGCGGAACGTGCTGCGCTTCGGAG 2735
          ||||||||| |  | |||  |||||  |||||||  | || |||  || || | ||
Db   1969 GACCCGCGCTCCTTCGGCGTCGCCGAACTCGGCCCGACGGGGGGCGTCGCGGCCTGGTG 2028

Qy   2736 GAGAAACCCGACGTCCCGCGCGCAGTTCGCTCGCGCCAGTTCATCGGCTGTGTACGGCCCG 2795
          ||||| |  || |||||| |  |||| | | || |  | | ||| |  | | | | |||
Db   2029 GAGAGCCGGTGCAGCGCCCCAGCACGCAGCGCCACTGGCCTGATCGGGGTCGATCGGCTACTTCTTCACCGCC 2088
```

```
Qy   2796 GCCGTCCAGGAGGCGGTACGGGCGGCCATCACCCCCCTCCGGCGCCGGGCGAGCTGGAGATCACC 2855
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   2089 GCCATCCACCAGGCGGTGGGCGGCCATCTGCCCAGCAGCTGCGCCGGCGGGAACTGGAGATCACC 2148

Qy   2856 CACGCGGCGCAGTGGATGATGAGCAGCCGGCGGTTGCGGCCGGAGACCACCACCACCCGG 2915
          |||||| ||||||||||||||||||||| |  |||| ||  || |||| || |||||
Db   2149 GACGCGCGTCCAGTGGTTGGTCACCTCCGGCCGCCGGCGACGTGCGCGCGTCCTTACGACGGC 2208

Qy   2916 CCCTGGGCGACACCGGCAGCGGGGAGACATGCTGGAGGTCAACCGTCACGTCCTGAC 2975
          |||| |  || || | || || ||  |  | |||  || | ||  || || |||||
Db   2209 TACTGGAAGGACACCGGCGAGGGTCGCAGAGCGTCCTTGAGTGTGCAACAGCCACCTCCTGAC 2268

Qy   2976 GGACTGAGGGCCGCATCGAGGGGAAGGTCGACGCGCACAGCAGCTGGTCGGCCGGGGTC 3035
          ||| |  || ||| ||| | | |||   |  || |  |||||||| |||| ||| |||
Db   2269 GGCCTGACCCGCGCGTCGACGGACAGTGGGGTCACACGTCGGTCGTCGTCGTCGGCCGGGGTC 2328

Qy   3036 CGGGTGCGCCGAAGGCGCCATCGTGCGGGGGTCACACGGTGGTGGCCCGGTGGTGATCGGC 3095
          | || | ||| ||||| || ||||| ||| |  ||| | ||| ||||||||| || |||
Db   2329 GTGATCGAGGGCGGGGGCCGCGCATCGTGCGGTCGCGGGGTCGAGGGCCCGATCATCGGC 2388

Qy   3096 GCGGGTGCCGTCGTCAGCAACTCCAGTGTCGGGCCCTGCACCTTCCATCGGGAGGACTGC 3155
          || ||||||| || | | || || |  | | |||  |  |  ||||| |||| |||||
Db   2389 GCGGGCACGGTCCTTCAGGACAGCCAGGTGGCCCGACCACACCTTCCATCGGGCGGGACTGC 2448

Qy   3156 CGGGTCGAGGACGAGCCGCCATCGAGTACTCCGTGCCTCGTGCCGCGGCGCCAGGTGCGAGGGGG 3215
          ||||| | | ||  || ||  |  | |||||||  |  ||| ||| |||  |  ||| | |||||
Db   2449 ACGGTGACGGAGCAGCGCCCGTCCATGCCCGGAGGGCTCGGCTCGGTCGCCGGAGGGCTCGGCCGCGGCC 2508

Qy   3216 GCGTCCCGCATCGAGGCGTCCCTCATGGCGCCGGCGCCCGTGCTGCGCGGCGGCTGGCCCCGT 3275
          | |||  ||  ||   | | |  |  ||| || ||||| |   ||| || ||||  |||||
Db   2509 GTGCGGCGGCCTGGCCGCAACTCGGTGATCGGGCGGCCCGTCGGCGCACCACCGGCGCCC 2568

Qy   3276 CTCCCGCAGGCTCACCGACTGGTGATCGCGACCAGCAAGTGTATCTCACCCCATGA 3335
          |  |  ||||| ||||||  |  || ||||||||| | | || | ||| ||||  ||||
Db   2569 GGCACGGCCCATCACTGCCTGGTCGTCGTGCGGAGACCACCGAGTGGAGGTCGCGGCATGA 2628
```

```
RESULT 6
US-08-681-953-31
; Sequence 31, Application US/08681953
; Patent No. 5710032
; GENERAL INFORMATION:
;   APPLICANT: Piepersberg, Wolfgang
;   APPLICANT: Stockmann, Michael
;   APPLICANT: Taleghani, Kamplz Mansouri
;   APPLICANT: Distler, Jurgen
;   APPLICANT: Grabley, Susanne
;   APPLICANT: Sichel, Petra
;   APPLICANT: Brau, Barbara
;   TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes
;   TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their
;   TITLE OF INVENTION: Use.
;   NUMBER OF SEQUENCES: 34
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;     ADDRESSEE: Dunner
;     STREET: 1300 I Street, N.W.
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: United States
;     ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/681,953
;     FILING DATE: 30-JUL-1996
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/196,218
;     FILING DATE: 25-AUG-1994
```

Qy  2916  CCCTGCCGCGACACCGGCAGCCGGAGGACAATGCTGGAGGTCAACCGTCACGTCCTGGAC  2975
          ||||| |||||||||  ||||||||  |||||  |||  |||  |||| |||||||||||||
Db  2209  TACTGGAAGGACACCGGGAGGGGTCGAGGACGTCCTTGAGTGCAACAGCACCTCCTGGAC  2268

Qy  2976  GGACTGGAGGGGCCGCATCGAGGGGAAGGTCGACGCGCACAGCACGTTGGTCGGCCGGGTC  3035
          ||| ||||    ||||||||   ||||| |||||||||  |||||||||||||||||||||
Db  2269  GGCCTGACCCCGCGCGGTCGACGGACAGGTCGACGCGCCGACAGCGTGCTCTCGGCCGGGTC  2328

Qy  3036  CGGGTGGCCGAAGGCGGATCGTGCGGGGTCACACGTGGTGGGCCCGTGGTGATCGGC  3095
          | |-| ||||||| |||||  |  ||||  |||| |-| ||||||  ||||||||||
Db  2329  GTGATCGGAGGCGGGGGGCGCATCGTGCGGGGTCCGGGGTCGAGGGCCCGGATCATCGGC  2388

Qy  3096  GCGGGTGCCGTCGTCAGCAACTCCAGTGTCGGCCCCGTACACCTCCATCGGGAGGACTGC  3155
          ||||| |-| |||  ||  ||| ||   || |||  ||  ||||||||||  |||||||||
Db  2389  GCGGGCACGGTCCTTCAGGACACCAGGTGGGGCCCCGCCACACCTCATCGGGCGGGACTGC  2448

Qy  3156  CGGGTCGAGGACAGCGCCATCGAGTACTCCGTCCTGCTCGGCGCCGGCGTCGTTGCCGGG  3215
          ||| ||||||   ||||||||   |||  |||||||  |||  ||||||||||  |||||||
Db  2449  ACGGTGACGGACAGCCGGCTGGAGGGCTCCATCGCCGGCGGCGCCGTCGTCGGTCACCGGC  2508

Qy  3216  GCGTCCCGATCGAGGGCGTCCCTCATCGGCGGCGGCGCCGGCGCCCCCGT  3275
          |-|-||||  ||||||  |||||||||||||  |||||||  ||||||||||
Db  2509  GTGCGGCCCTGCCAACTCGCTGATCGGCGGGGTGCTGATCGGCGGCACCGGCCCCC  2568

Qy  3276  CTCCCGCAGGCTCACCGACTGGTGATCGGCGACCACGCAAGGTGTATCTCACCCCATGA  3335
          ||     ||  ||  |||||||  |||| ||||||||||  |||||| ||||||||||
Db  2569  GGCACGGCGCATCACTCACTGCCTGGTCTGGTCGTCGGAGACCACCACCGGAGTGCGGCCATGA  2628

RESULT 7
US-09-036-987A-25
; Sequence 25, Application US/09036987A
; Patent No. 6143526
GENERAL INFORMATION:
-  APPLICANT:   Baltz, Richard H.
-  APPLICANT:   Broughton, Mary C.
-  APPLICANT:   Crawford, Kathryn P.
-  APPLICANT:   Madduri, Krishnamurthy
-  APPLICANT:   Merlo, Donald J.
-  APPLICANT:   Treadway, Patti J.
-  APPLICANT:   Turner, Jan R.
-  APPLICANT:   Waldron, Clive
-  TITLE OF INVENTION:   Biosynthetic Genes For Spinosyn Insecticide
-  TITLE OF INVENTION:   Production
-  NUMBER OF SEQUENCES:   39
-  CORRESPONDENCE ADDRESS:
-     ADDRESSEE:  Dow AgroSciences LLC Patent Department
-     STREET:  9330 Zionsville Road
-     CITY:  Indianapolis
-     STATE:  Indiana
-     COUNTRY:  USA
-     ZIP:  46268
-  COMPUTER READABLE FORM:
-     MEDIUM TYPE:  Floppy disk
-     COMPUTER:  IBM PC compatible
-     OPERATING SYSTEM:  PC-DOS/MS-DOS
-     SOFTWARE: -PatentIn Release #1.0, Version #1.30
-  CURRENT APPLICATION DATA:
-     APPLICATION NUMBER:  US/09/036,987A
-     FILING DATE:  09-MAR-1998
-     CLASSIFICATION:  435
-  ATTORNEY/AGENT INFORMATION:
-     NAME:  Stuart, Donald R
-     REGISTRATION NUMBER:  28,479
-     REFERENCE/DOCKET NUMBER:  50,608
-  TELECOMMUNICATION INFORMATION:
-     TELEPHONE:  (317)337-4816
-     TELEFAX:  (317)337-4847
INFORMATION FOR SEQ ID NO:  25:
-  SEQUENCE CHARACTERISTICS:
-     LENGTH:  2310 base pairs
-     TYPE:  nucleic acid
-     STRANDEDNESS:  double

ATTORNEY/AGENT INFORMATION:
-  NAME:  Ogden, Stasia L.
-  REGISTRATION NUMBER:  36,228
-  REFERENCE/DOCKET NUMBER:  02481.1372-00000
-  TELECOMMUNICATION INFORMATION:
-     TELEPHONE:  202-408-4000
-     TELEFAX:  202-408-4400
INFORMATION FOR SEQ ID NO:  31:
-  SEQUENCE CHARACTERISTICS:
-     LENGTH:  2634 base pairs
-     TYPE:  nucleic acid
-     STRANDEDNESS:  single
-     TOPOLOGY:  linear
-  MOLECULE TYPE:  DNA (genomic)
-  FEATURE:
-     NAME/KEY:  CDS
-     LOCATION:  3..401
-  FEATURE:
-     NAME/KEY:  CDS
-     LOCATION:  416..1531
-  FEATURE:
-     NAME/KEY:  CDS
-     LOCATION:  1561..2625
US-08-681-953-31

Query Match             7.3%;    Score 497.6;  DB 1;  Length 2634;
Best Local Similarity  66.3%;    Pred. No. 8e-64;
  Matches  716;  Conservative    0;  Mismatches  364;  Indels   0;  Gaps    0;

Qy  2256  GCGGAGGCCCGAGGTGAAGGCCCTGCTCCTGGAACCGGCAGCAGACTGAGGCCG  2315
          | |-| | |     ||||||| |  | |  | || || || || || | ||| ||||
Db  1549  GAGGACCTCCGATGAAGGCTCTGGTGCTCGGCGGATCTGTACCCGCCTGCGGCCT  1608

Qy  2316  TTCACCCACACCGGCCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTAC  2375
          ||||| || |||| |||  || ||||| ||||| ||||  |||||| ||||||||||||
Db  1609  TTCAGTTATTCGATGCCCCAAACAACTGATCCCCATCGCCAACACACCCGCTGCTGGTGCAT  1668

Qy  2376  GCGCTGGAGTCCCTCGCCGGCGGGGTCCGGGAGGCCGGCGTCGTGTGGGCGCGTAC  2435
          | ||||-| ||-||  |||| ||  |||| | ||| |||||||  |||||||||||
Db  1669  GTGCTGAACGCCGTCGGGGAGCGTGGGCGTGACCGAGGTCGGCGTCATCGTCGGCAACCGC  1728

Qy  2436  GCCCGGGAGATCCGCGAACTCACCGGCCGCCACCGCGTTCGGGTTACGCATCACCTAC  2495
          |||| ||||||  | |||||||||| | |-| |||| |  |||| || | ||||||||
Db  1729  GGCCCCGGAGATCGAGGGCGTGCTCGGACGACGGTGCCCGGTTCGACGTGGCGCATCACCTAC  1788

Qy  2496  CTCCACCAGCCCCGCCCGCTCGGTCTCGGCCACGCGGTGCGATTGCGCCGCGGCTTCCTG  2555
          ||| ||||   ||| ||  ||  || || |||| |  | | ||||||  |||| ||| |||
Db  1789  ATCCCCCAGGACGACGCACCGCGGCACCGGACTGGCCCACACCGTGTCCATCGCCCCGGCGGCTTCCTC  1848

Qy  2556  GGCGACGACGACTTCCTGCTGTACCTGGGGGACAACTACCTGCCCCAGGGCGGTCACCGAC  2615
          |||||||||||||||||||  |-|||||||| |||| || | ||| |||  ||||||||||
Db  1849  GGCGACGACGACTTCGTGATGTACCTCGGCGACAACATGCTGCCCGACGGAGTCACCGAG  1908

Qy  2616  TTCGCCCGCCAATCGGCCCGCCGATCCCGGCGTCCCGGCGCCGCTGCTCACCCCGGTCGCG  2675
          ||||| || |||     |||||  ||   | |-| ||| |||||||| ||  ||||| |||
Db  1909  ATCGCCGAGGAGTTCACCCGGCAGCCCCGCCCCCCCAGGTCGTCGTGCACAAGGTCCCC  1968

Qy  2676  GACCCGGTCCGCCTTCGGCGTCGGAGGTCGAGCGGACGGGAACGTGCTGCCTTGGAG  2735
          ||||| |-|||||| | | ||| |||  ||  | |||||||||  || || |||||
Db  1969  GACCCGGGGTCCTTCGGCGTCGGGCGAACTCGGCCGACGAGGTGCTGCGCCTGGTG  2028

Qy  2736  GAGAAACCCGACGTCCGCGCAGCTCGCTCGCGACGCCGTGTACGCCCG  2795
          |||| || |        ||||||||  |  | |  |||-| || |||||||
Db  2029  GAGAAGCCGTGGCAGCCGCAGCCGCAGCGGCACATGGCCCGCTGTTCGTCTACTTCTTCACCGCC  2088

Qy  2796  GCCGTCCACGAGGCGGTACGGGCCCATCACCCCTCCGCGGAGATCACC  2855
          |||||||||||||| || ||||||||  |||||  |||||||||||
Db  2089  GCCATCCACCACGGCCGTGGCGCGCCCATCTCCCCCAGCAGCCCGGCGAACTGGAGGATCACC  2148

Qy  2856  CACGCCGTGCAGTGGATGATGATCGAGCCGGGCTTCGCCGTGCGGTACGGGCCGAGGCCACCCGG  2915
          |||||||| |||||  |  || || | |  ||| |||| ||||||||||  |-|
Db  2149  GACGCCGTTCCAGTGGTTGGTCACCTCCGCGGACGTGCGCGCACCGGGCGCGGCGTGCCTCTACGACGGC  2208

```
;
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 88..1077
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1165..1992
US-09-036-987A-25

Query Match       6.8%;    Score 465.8;   DB 3;   Length 2310;
Best Local Similarity 68.5%;  Pred. No. 2.9e-59;
Matches 660;  Conservative 0;  Mismatches 297;  Indels 6;  Gaps 1;

Qy  3344  ATCCTCGTCACCGGCGGAGCGGGCTTCATTCGCTCCGCCTACGTCCGGCTCCTGTCG      3403
          ||| ||||||||||||||||| ||||||||| |||||||||||| ||||  ||| || ||
Db    94  ATTCTGGTCACCGGCGGAGCCGGGCTTCATCGCTCCGCCTACGTTCGGCAGTTGCTCGGT  153

Qy  3404  CCCGGGGCCCCCGG------CGGCGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCC  3457
          ||  |  ||| |          ||||| ||| ||||| ||  | || ||||||||||||||
Db   154  GGTGCGTACCCCGCATTCGCCGACGTGGTCGTCGACAAGCTCACCTACGCC           213

Qy  3458  GGCAGCCTCGCCGCCTGCACGCCGGTGCGTGACCATCCCGGCCTCGACCTTCGTCCAGGGC  3517
          |||| ||  ||   ||| || ||||| ||||| || ||| |||||||||||||||||| |||
Db   214  GGCAACGAGGCGAACCTGCCGCCGGTCGGCGACAACCCCGGCTGAAGTTCGTCGTCGCGGC  273

Qy  3518  GACGTGTCGCGACACCGCTCGTCGACACGCTGGCCGCGCACGACGCAATCGTCGCTGCGCAC  3577
          |||| | ||| || |||  ||||  |||| |||||||||||||| |||  ||| || ||||||
Db   274  GACATCTGCGACACCGCCGCGAACTGGTTCATCACCGCGTCGGACGTGGTGGACGTGGTCGAC  333

Qy  3578  TTCGGCGGCCGAGTCGCACGTCGACCGCTCCATCACCGACAGCGGTGCCTTCACCCGCACC  3637
          ||||||||||| |||||||| ||| |||| || ||  | ||||||| ||  |||| |||||
Db   334  TTCGGCGGCCGTACCCGCCATTCGCCCGCTGTTCATCACCGACGTGGTCGTTCGTGACGGGC  393

Qy  3638  AACGTGCTGGCCACCCAGTCCTCCGCACGGCCGCTCCGCGACCGGTGCGCCGCACCTTC  3697
          |||||||||||| |||||  ||| || || |  |  ||   ||||| || || |||||||
Db   394  AACGTGGTCGGCCACCAACGTCTGCTCAGCGTGCTCGACGCCGAGATCGGCAAGTTC  453

Qy  3698  GTGCACGTCTCCACCGACGAGTCTACGGCTCGGGCGCCCGCCGCGAGAGC  3757
          |||||||||||||||||||||||| || || ||||| |||| | || ||
Db   454  GTGCACGTGGTCGCGGAACTGCGAGGACGTCTACGGCTCACGGGGCCGAAGAC  513

Qy  3758  GACCCCTGCTTCCGACCTCGCCGTACGGCGGCGTCGAAGGCGTCGAAGCCATCATGGCG  3817
          || |  |||   | ||  ||||| || | |||| || ||| || ||  ||| || ||
Db   514  CACGCCGTGGAGCCGAATTCCCGTACTCCGGCGGCGAAAGCGGCTCGGACGTCGGACCTGGCC  573
                                           ... Gaps ...

Qy  3818  CTCGCCACCACCGGCACCGACGGGCTCGGGTGACCGTGTTCGAAACAACTTC  3877
          || |  || ||  ||  || |||| || |  ||||| |||||| | |||||||
Db   574  CGGCCTACCACCGGGCATCCGCACCGGCACTGCCGTGTGCATCACCCGCTCCAACAACTAC  633

Qy  3878  GGCCCCCACCAGCATCCCGAGAAGCTCATACCGGCCCTTCCTGACCAGCCTCCGTCCGGC  3937
          |||||| |||| ||  |||||| | ||| |  |||| |||| |||||||||| |  |||||
Db   634  GGCGGCCTACCAGTTCCCGGAGAAGGTGCTGTCGCCGTGTCTTCATCAACCTGATGACGGGC  693

Qy  3938  GGCACCGTTCCCCTCTACGGCGACGGGCCGCACTGCGCAGTCGCTGCGACGAC  3997
          |||  |||| ||  ||||| || || |||| || ||| || | |||| ||||
Db   694  AGCCAGGTGCGCGTCGCCTCAACCGTCGCGCACGTGGGACTGCGCACGTCAGCGAC  753

Qy  3998  CACGTCAGGCGCGCAACTCGTCCGGGTCGGCGGCCGGAGAGATCTACAACATC  4057
          || | |  ||| || || ||||| || ||||||  ||| | |||  ||| |||
Db   754  CACTGCCGGGGCATCCGGCCGCTCCGGGACTCCGCGGCGCGGGGATCTCTACAACATC  813

Qy  4058  GGGGCGGCCCCGGAGCTCACGGCACCTCGCCCAACCTGGAGCTCACGTGTCGCGGC  4117
          || || |  |||| | || || || ||  ||| |||||||| ||| | |||||||
Db   814  GCCGGCGGCAGCGAGCTGACCAACGAGCTGGTCTGGGCAGAGCTTCGTGGACTGGGC  873

Qy  4118  GCGGGCGCCCGGAGCGCCATCGTCCACGTCCGAGAAACCGCAAGGGCCACGACGCGCTACGCG  4177
          || ||| || ||  || ||  ||   || || ||   || ||  | || |  || || |||||
Db   874  CTCGACTGGTCGGTGTCGCGCGGTGTCGGCGGACACCCCCGCCCGGACACCCCGCGCTACTCG  933

Qy  4178  GTCGACCACCAAGAATCACCGCCGAACTCGGTTACCGCGACGACTTCCGACC  4237
          |||||||||||||||| || |  || |||  || ||  | |||||| | |||
Db   934  GTGCACCAGCAAGATCGTCGAGGAACTGGGGTACGCGCGACGAGTCTCAGGGAGAC  993
```

```
Qy  4238  GCGCTGGCCGACACCGCGAAGTGTACGAGCGGCCACGAGGACTGGTGGCGTCCCCTGCTC  4297
          ||| |||  || ||||||||||  ||||||   |||||||| |||||||| |||||||||
Db   994  GGGCTGCCGGAGCAACATCCGCTGGTACCAGACAACCGGGACTGGTGGAGCCGGCTGAAG  1053

Qy  4298  GCC  4300
          |||
Db  1054  GCC  1056
```

```
RESULT 8
US-09-370-700-25
;   Sequence 25, Application US/09370700
;   Patent No. 6274350
;   GENERAL INFORMATION:
;     APPLICANT: Baltz, Richard H
;     APPLICANT: Broughton, Mary C
;     APPLICANT: Crawford, Kathryn P
;     APPLICANT: Madduri, Krishnamurthy
;     APPLICANT: Treadway, Patti J
;     APPLICANT: Turner, Jan R
;     APPLICANT: Waldron, Clive
;   TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;   FILE REFERENCE: 50489 DIV1
;   CURRENT APPLICATION NUMBER: US/09/370,700
;   CURRENT FILING DATE: 1999-08-09
;   EARLIER APPLICATION NUMBER: US 09/36987
;   EARLIER FILING DATE: 1998-03-09
;   NUMBER OF SEQ ID NOS: 39
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 2310
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (88)..(1077)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (1165)..(1992)
US-09-370-700-25

Query Match       6.8%;    Score 465.8;   DB 4;   Length 2310;
Best Local Similarity 68.5%;  Pred. No. 2.9e-59;
Matches 660;  Conservative 0;  Mismatches 297;  Indels 6;  Gaps 1;

Qy  3344  ATCCTCGTCACCGGCGGAGCGGGCTTCATTCGCTCCGCCTACGTCCGGCTCCTGTCG      3403
          ||| ||||||||||||||||| ||||||||| |||||||||||| ||||  ||| || ||
Db    94  ATTCTGGTCACCGGCGGAGCCGGGCTTCATCGCTCCGCCTACGTTCGGCAGTTGCTCGGT  153

Qy  3404  CCCGGGGCCCCCGG------CGGCGTCGCGGTGACCGTCCTCGACAAACTCACCTACGCC  3457
          ||  |  ||| |          ||||| ||| ||||| ||  | || ||||||||||||||
Db   154  GGTGCGTACCCCGCATTCGCCGACGTGGTCGTCGACAAGCTCACCTACGCC           213

Qy  3458  GGCAGCCTCGCCGCCTGCACGCCGGTGCGTGACCATCCCGGCCTCGACCTTCGTCCAGGGC  3517
          |||| ||  ||   ||| || ||||| ||||| || ||| |||||||||||||||||| |||
Db   214  GGCAACGAGGCGAACCTGCCGCCGGTCGGCGACAACCCCGGCTGAAGTTCGTCGTCGCGGC  273

Qy  3518  GACGTGTCGCGACACCGCTCGTCGACACGCTGGCCGCGCACGACGCAATCGTCGCTGCGCAC  3577
          |||| | ||| || |||  ||||  |||| |||||||||||||| |||  ||| || ||||||
Db   274  GACATCTGCGACACCGCCGCGAACTGGTTCATCACCGCGTCGGACGTGGTGGACGTGGTCGAC  333

Qy  3578  TTCGGCGGCCGAGTCGCACGTCGACCGCTCCATCACCGACAGCGGTGCCTTCACCCGCACC  3637
          ||||||||||| |||||||| ||| |||| || ||  | ||||||| ||  |||| |||||
Db   334  TTCGGCGGCCGTACCCGCCATTCGCCCGCTGTTCATCACCGACGTGGTCGTTCGTGACGGGC  393

Qy  3638  AACGTGCTGGCCACCCAGTCCTCCGCACGGCCGCTCCGCGACCGGTGCGCCGCACCTTC  3697
          |||||||||||| |||||  ||| || || |  |  ||   ||||| || || |||||||
Db   394  AACGTGGTCGGCCACCAACGTCTGCTCAGCGTGCTCGACGCCGAGATCGGCAAGTTC  453

Qy  3698  GTGCACGTCTCCACCGACGAGTCTACGGCTCGGGCGCCCGCCGCGAGAGC  3757
          |||||||||||||||||||||||| || || ||||| |||| | || ||
Db   454  GTGCACGTGGTCGCGGAACTGCGAGGACGTCTACGGCTCACGGGGCCGAAGAC  513
```

```
Db  1354  GGGCGTACCCGACGTCCCGCCGGTGCCCGCCGGTGAGGTGACGTGATCGTCCTGGACAGCCTCACCTTCGTCCAGGGCG  1295
Qy  3459  GCAGCCTCGGAGCCGAATTCCCGGACTCGGCGGAAAGCGGGCTCCGGACCTGCTGGCC  573
Db  1294  GCAACCGGCGCCAACCTGGCCCGGTGGAAGCGGACCCGGCGCGCGGCTGCGGGTTCGTCCACGGGCG  1235
Qy  3519  ACGTGTGCGACACCGCGTCGTCGACACGCTGGGCGCGTCCGCGGCACGACACATCGTGCACT  3578
Db  1234  ACATCCGGACGCGGCGGAACTCGCGCGTCGCACGCGTCGCGACTAC  1175
Qy  3579  TCGCGGCCGAGTCGCACGTCGACCGTCCATCACCGACGAGGTGCCTCACCGCCACCA  3638
Db  1174  TCGGCCCGGAGAGCCACGTCGACCGCTCCATCGCGCGGCGTCCGCCACCGGGACCGGGACCA  1115
Qy  3639  ACGTGCTGGGCACCCAGTCCTGCGACGTGGCTCGACGTGCGCACCTTCG  3698
Db  1114  ACGTCCAGGGCACGCAGACCCTGCTCCAGTCGGTCTCGACGCCGCGGGGCGTCGGGGTCG  1055
Qy  3699  TGCACGTCTTCACCGACGAGGTGTACGGCGTCGAAGGGCGTCGGACCTCATGGGCC  3758
Db  1054  TGCACGTCTTCACCGACGAGGTGTACGGCGTACGGGCGTCGATCGACTCGGTGAACGAGGAGCA  995
Qy  3759  ACCCCTGTTCCGACCTCGCCGTACGGCCGTGCCTCGCTCAAGGCCGGCTCCGGAGCAACTTCG  3818
Db  994   GCCCGCTGGAACCCAACTCCCCTACGCGGCCTCGGGTGACCTGCTTCCGCCC  935
Qy  3819  TGGCCCACCACGCACCCAGGCCTGGACGTCCGACGTCACCGGATCACGCGCTTCGTGTCCGAACAACTTCG  3878
Db  934   GGGCCTACCACGCCACCTACGGCGTCGACGTACGGATCACCCGGCTGCTGCAACAACTACG  875
Qy  3879  GCCCCCACCAGCATCCGAGAAGTCTCATCACCGGCTTCGACCAGCCTCCGTGTTCCGGGCG  3938
Db  874   GGGCCTACCACGCACCCCGGAGAGCTCATCCCCTCCTTCGTGACGAACGTCCTCGACGGGCG  815
Qy  3939  GCACCGTTCCCCTCTACGGACGGACGGCGGCGCGCGTGCGCGCTGCGCACGTCGACGACC  3998
Db  814   GGACCGCTCCCCGTGTACGGCGACGTCGGCGCCGAGTGGTCCCGGAGGAGATCTACAAACATCG  755
Qy  3999  ACGTCAGGGCGCGTGCGAACTCGTCCGCGTGTCGGCGGCCGGGCGGCGGCGAGGTCTACCCATCG  4058
Db  754   ACTGCGGGGCGATCGGCGATCCGAACTCCTCCGGCGCCGGTTGCTCGCCCACTGTGCGCGCGG  695
Qy  4059  GGGGGCGGCCACCTCGCTGCGCCAACCTGGAGCTGCGGCACGCGGTGCGCACCGGTGCTCGCCACTGTGCGCGCGG  4118
Db  694   GCGGCGGCCGGCGCTGCGACTGCGGCGTCGAGCTGGAAGTGCGGGTGAAGTGGTCTGGCCGCGCTCGGGCGG  635
Qy  4119  CGGGGCCGGAGCGATGGCCATCGTCCACGTCGAGAACGGCAAGGGGCACCCAAGGGGCTACGCGG  4178
Db  634   CCGACTGGTCCTCGGTCGTCCGGCGGTCCCACCCCCCAGGCGCCACGGCCACCGCGCTGCGCTACTCCC  575
Qy  4179  TGACCACAGCAAGATCACCGCGACATCGGTTACCGGCGCCGACTTCCGGCACCG  4238
Db  574   TGCACGCGGCCAAGATCGACGCGACTCCGCGCTACCGCGCGCACCGCAGATCGCGCTTCGCGGACG  515
Qy  4239  CGCTGCGCGGCCGAAGTGGTACGACGCGCACGAGGACTGGTGGTGCGGGACCGTCCCCTGCTCG  4298
Db  514   GCCTCGCGCGCCACCATCCCTGGTTACCCGGCGGGAACCCCGGCCCGCCCCCACCGC  455
Qy  4299  CCGGGCACATGACGTCGGGGCCGGCCGGGACCCAACCACCCGGCCCGCCCACCGC  4347
Db  454   CCACCCGCGCGCCAGCCCGCCGCCCCACCGCCGCCCCCCGCCCCAGCGCCGCCGCCCACCGC  406
```

RESULT 10
US-09-105-537-13
; Sequence 13, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.

```
Qy  3758  GACCCCTGCTTCCGACCTCGCCGCGGGCTCGGAAGGGCGGCTCGAAGGCGCCGTACGGCGGCGTCGGACCTTCATGGCG  3817
Db  514   CACCGCGTCGGAGCCGAATTCCCGGGACTCGGCGGCGAAAGCGGGCTCCGGACCTGCTGGCC  573
Qy  3818  CTGCCCACCACGCACCCCACCCACCGGACTCGGCGGTGACCCGCGTGTTCGAACAACTTC  3877
Db  574   CGCCGCTACCACGCCACCACCGGACTCGGTGCATCACCCGCTGCTGCAACAACTAC  633
Qy  3878  GGCCCCACCACCAGCATCCCGAGAAGTCATACCGCGCCTTCGACCAGCCTCCTGTCCGG  3937
Db  634   GGGCCTACCAGTTCCCGGAGAGGTGCTGCCGCTGTCATCATCGACGACTGATGGACGGC  693
Qy  3938  GGCACCGTTCCCCTCTACGGACGGACGGCGGCGGCGCACCTGGCTGCCGACGTGCACGAC  3997
Db  694   AGCCAGGTGCCGCTCTACGGACGGCGGCGTCCAACCTGCGGGACTGGCTGCACGTGCACGGAC  753
Qy  3998  CACGTCAGGGCCGTCGACTCGTGCGGCCCGGGCGCCGGCGAGAGATCTACAAACATC  4057
Db  754   CACTGCCCGGGCGCACCTGGCATCGAGCGCATCCAGCTCTGGGGCAGCCTGTGCGGC  813
Qy  4058  GGGGGCGGCCACCTCGCTGCGCCAACCTGGAGCTGCGACGTGTGGCGGCGCTACCG  4117
Db  814   GGGCGGCCGGCGCTGCGACCTGCGGTGTCGCGCGGCGGCGACACGCCGCGCGACTGCGGC  873
Qy  4118  GCGGGGCCGGAGCGCATCGTCCACGTCGAGAACGGCCAAGGGGCCACGCGGCGCTACCG  4177
Db  874   CTCGACTGGTCGGTGGTGCGGCGGCCGTCACCGACCGGCGTCGACCTTCGGTACCCGCCGACC  933
Qy  4178  GTCACCACAGCAAGATCACCGCGCACATCGGTTACCGCCGCCGCCGCACTTCGCGACC  4237
Db  934   CCGACTGGTCGTCGGTGGTCCGGCGGTCCCACCCCCCAGGCGCCACCGCGACTTGCGCC  993
Qy  4238  GCGCTGGCCGGACACCGGCGAAGTGTAGAGCGGCGCAGGAGGACTGGTGGTCGCGTCCCCTGCTC  4297
Db  994   GGGCTGCGCGAGGACAAATCCGCTGCTGTACCAGGACAACCCCGGACTGGTGGGAGCCGCGCTGAAG  1053
Qy  4298  GCC  4300
Db  1054  GCC  1056
```

RESULT 9
US-09-434-288-12/c
; Sequence 12, Application US/09434288
; Patent No. 6303767
; GENERAL INFORMATION:
; APPLICANT: Betlach C., Melanie
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20030.00
; CURRENT APPLICATION NUMBER: US/09/434,288
; CURRENT FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/107,093
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Streptomyces narbonensis
US-09-434-288-12

Query Match       6.6%;   Score 454.2;   DB 4;   Length 1476;
Best Local Similarity 66.4%;
Matches  670;   Conservative   0;   Mismatches 333;   Indels   6;   Gaps   1;

```
Qy  3345  TCCTCGTCACGGCGGAGCGGGCTTCATTCGCTCGCCTACGTCCGGCGGCTCCTGTCGC  3404
Db  1414  TTCTGGTGACCGGAGGTGCGGGGCTTCATCGGTGCGGCGGCTCCTCGCCG  1355
Qy  3405  CCGGGGCCCCCGGCGGCGGCCGTGCGC-----GGTGACCGGTCCTCGACAAACTCACCTACCGCCG  3458
```

```
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
;   LENGTH: 1014
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-13

Query Match             6.5%;  Score 447;  DB 4;  Length 1014;
Best Local Similarity  66.8%;  Pred. No. 1.5e-56;
Matches  655; Conservative   0; Mismatches 320; Indels   6; Gaps   1;

Qy   3345 TCCCGTCACCGGCGGAGCGGGCTTCATTCGCTCGCCCTACGTCCGCGCTCCTGTGTCGC 3404
               ||-|| |||| || |||||||||||  |||  ||||||| ||||| |||||||
Db      8 TTCTGGTGACCGGAGGTGCGGGCTTCATCGCTCGCACTTCGTTCGGCAGCTCCTCGCCG  67

Qy   3405 CCGGGGCCCCCGGCGCGGCGTCGC------GGTGACCGGTCCTCGACAAACTCACCTACGCCG 3458
               ||-|  |||| || |||||| |||       |||||||||||| ||||||||||||||||
Db     68 GGGCGGTACCCGGACGTGCCCGCCGCCGATGAGGTGATCGTTCCTGGACAGCCTCACCTACGCGG 127

Qy   3459 GCAGCCTCGCCCGCTGGACGCGGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCG 3518
               ||||||||||||| || ||||  |||| |||||||||||||||  |||||||||||||
Db    128 GCAACCGGCCAACCTCGCCCCGGTGGACGCGGACCCGGACTGGCCTTCGTCCACGGCG 187

Qy   3519 ACGTGTGCGACACCGCCGTCGTCGACACCGTGGCCGCGCGACGACATCGTGCACT 3578
               ||-| |||||| | ||||||  |||| |||||  |||||||||||||||
Db    188 ACATCCGGCAGCGCCGGCCTCCTCGCCCGGGAACTGCGCGCGTGGACGCCATCGTCCACT 247

Qy   3579 TCGCGGCCCAGTCGCACGTCGACCGCCTCCATCACCGACACGCGGTTGCCTTCACCCCCACCA 3638
               ||||||||||||||||||| |||||||||||||  ||||  ||||||||||||||||||||
Db    248 TCGGGGCCGAGAGCCACGTGGACCGCCTCCATCGCGGCGGCCTCCGTGTTCACGGAGAGACCA 307

Qy   3639 ACGTGCTGGCACCCAGTCCTCGCTCGAGCGCGGCCCGCTCCGCCACGGTGTCGCACCTTCG 3698
               |||||||||||||||| ||| |||  |||||  ||||||||||||||||||||||||||||
Db    308 ACGTGCAGGGCACCAGCACCAGAGCGTCCAGTGCGCGCCGTGCCGCGCCCGGTGCGGGGTCG 367

Qy   3699 TGCACGTCCACCGACGAGGTGTACGGTCCGGTCGGTGATGACTCCGGCTCCTGGACCGGAGAGCA 3758
               ||||||||||||||||  ||||||||||| || ||| |||  | ||||| ||||||||||||
Db    368 TGCACGTCTCCACCGACGAGGTGTACGGGTCGATCGGCCGGTCCGGGTCCTGGACCGGAGACGA 427

Qy   3759 ACCCCCTGCTTCCGACCTTCCGACGCCGGCGTACCGGCGTCGAAGGCGGCCCATGCGCGC 3818
               ||-||| |||  ||||| |||| |||||||||| ||| |||  |||||||||||||
Db    428 GCCCGTGAGCCCAACTCGCCCCTACCGGCCGTCCAAGGCCGGCTCCGACCTCGTTGCCC 487

Qy   3819 TCGGCCCACCACCGGCCCACCACCCGGTTGCTCGGGTGACTTCGGGTGTTCGAACAACTTCG 3878
               ||||| |||||||||| ||  || ||||  |||||||||||||  |||| ||||||||||
Db    488 GCGGCCTACCACCGGCGTACGGCCTCCGACGTACCGGATCCATCCCGGCTGCTGCAACAACTACG 547

Qy   3879 GCCCCACCAGCAGCATCCCGAGAAGCTCATACCCGGCTCGTCCCGTCCGTCCGGGCG 3938
               ||||||||||||| |||||||||||||  ||| |||||| ||||  |||||||||||||
Db    548 GGCCGTACCAGCAGCACCCCGAGAAGCTCATCCCGCTCCTCGACGGCCTGACGCGGCG 607

Qy   3939 GCACCGGTTCCCCTACGGGCGACGGGCCATCGCCGACGTGCCGCACGTCGACGACC 3998
               || |||||| ||||| |||  ||||||||||||||| ||  |||||||  |||||||
Db    608 GGACTGGTTCTCGGTACGGGGCGACCTCGGCCCACGGCCCGAGTGGTCACACCGACGACC 667

Qy   3999 ACGTCAGGGCGGTCGAACTCGTCCGCGTGTCCGGCCGGCGGCGGGAGAGATTCACAAACATCG 4058
               |||| ||||||| || |||||||  ||| || |||  ||||||||||||||| ||||| |
Db    668 ACTGCCGGGCATCGCGTCGCTCCTCCGCGGCCGGCGGGCCGGAGATCGTTACCACCATCG 727

Qy   4059 GGGGCGGCACCTGCTGCTCCCAACCTGGAGCTCACGCACCGTTGCTCGACGTGCGGCG 4118
               |||||| ||| || |||  |||| ||||  ||| |||| |||  ||||||||||  || |||
Db    728 GCGCGGGCCTGGAGCTGCACCAACCCGGCGATCCTCCTGGACTCCGCTCGGCGGCTGCCG 787

Qy   4119 CGGGGCCCGAGGCGCCATCGTCCACCTCGAGAACCCGAGACCCGGCGCTACGCGG 4178
               || ||||||||||| || ||  ||||||||||||||||||   |||||||||||
Db    788 CCGACTGGTTCTCGGTCCGGAAGTCGCCGACCTCGCGGCCTACTCCC 847

Qy   4179 TCGACCAGCAGCAAGATCACCGCGGCAAGCTCGGTTACCGGCCGCCACCGACTTCGGCGACCG 4238
               |||||  || |||||||||||||||||||||||||||  |||||||||||| |||||
Db   2089 CCGGCACCCAGGTGACCGACGACCATATCAGCAAGCCCCGTAAATGGGCTGGCCGAACATCGGGCA 2030
```

```
Db    848 TCGACGGCGGCGAGATCGAGCGCGAGCTCGGCTACCGCCCCAGGGTCTCCTTCGCGGACG  907
               ||||| ||| |||| ||||  |||||||||||||||||||||||||||||
Qy   4239 CGCTGCCCGACACCGGCGAAGTGGTACGACGGCACGAGGACGACTGGTGCCGCTCCCTGCTCG 4298

Db    908 GCCTCGCCGCGGACCGTCCGCTGCTCGTACCGGGAGAACCCGGGCGGCTGGTGGAGCCCGGCCTCAAGG 967
               ||||  ||||||||||||  |||  ||| ||||| ||||||||||||||
Qy   4299 CCGGGACATGACGTCGGGCCG 4319

Db    968 CGACCGCCCCCGCAGCTGCCCG 988
```

```
RESULT 11
US-08-606-322-1/c
; Sequence 1, Application US/08606322
; Patent No. 5753501
; GENERAL INFORMATION:
; APPLICANT:  Crueger, Anneliese; Piepersberg,
; APPLICANT:  Wolfgang; Distler, Jurgen; and
; APPLICANT:  Stratmann, Ansgar
; TITLE OF INVENTION:  ACARBOSE BIOSYNTHESIS GENES FROM
; TITLE OF INVENTION:  ACTINOPLANES sp., PROCESS FOR THE ISOLATION
; TITLE OF INVENTION:  THEREOF AND THE USE THEREOF
; NUMBER OF SEQUENCES:  5
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;   STREET:  660 White Plains Road
;   CITY:  Tarrytown
;   STATE:  New York
;   COUNTRY:  U.S.A.
;   ZIP:  10591-5144
; COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 mb
;   MEDIUM TYPE:  storage
;   COMPUTER:  Bravo 3/25s
;   OPERATING SYSTEM:  DOS
;   SOFTWARE:  WordPerfect 5.1
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/606,322
;   FILING DATE:  23-FEB-1996
;   CLASSIFICATION:  435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  DE 195 07 214.6
;   FILING DATE:  02-MAR-1995
; ATTORNEY/AGENT INFORMATION:
;   NAME:  Kurt G. Briscoe
;   REGISTRATION NUMBER:  33,141
;   REFERENCE/DOCKET NUMBER:  Bayer 9537-KGB
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (914) 332-1700
;   TELEFAX:  (914) 332-1844
;   TELEX:
; INFORMATION FOR SEQ ID NO:  1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH:  2219 Nucleotides
;   TYPE:  Nucleic Acid
;   STRANDEDNESS:  Double
;   TOPOLOGY:  Linear
;   MOLECULE TYPE:  Genomic DNA
;   HYPOTHETICAL:  No
;   ANTI-SENSE:  No
; ORIGINAL SOURCE:
;   ORGANISM:  Actinoplanes sp. SE 50/110
US-08-606-322-1

Query Match             5.7%;  Score 389.2;  DB 1;  Length 2219;
Best Local Similarity  57.6%;  Pred. No. 2.9e-48;
Matches  760; Conservative   0; Mismatches 548; Indels  12; Gaps   3;

Qy   3248 CGGCGCCGTCGTCGGCCCGGCCCCCCGTCTCCCGCAGGCTCACCGACTGGTGATCGGCGCA 3307
               ||||| |||||||||||  ||||||| ||||||||||||||||||||| |||||| ||||
Db   2089 CCGGCACCAGTGACCGACGACCATATCAGCAAGCCCCGTAAATGGGCTGGCCGAACATCGGGCA 2030
```

```
Qy   3308 CCACAGCAAGGTGTATCTCACCCCATGACCACGACCATCCTCGTCGCGGGGGAGCGGGGC 3367
          ||   | -|        ||||  | - |||       |||||||||||| |||
Db   2029 CCCATTGGCGGCATTGCTACAATCCCGGCCTCCTGTACCGGCCCGGGGCCCCCGGC 1970

Qy   3368 TTCATTCCTCCG--------CCTACGTCCGCGCTCCTGTCGTCCGGCCCTCCGGC 3418
          ||  |  |||   |||||    | |||||  ||||
Db   1969 TTTATCGGGTCCCATTTTGTAACTTCCCTGATCAGTGGCGACATTGCCACCACACCC 1910

Qy   3419 GGCGTCCGCGTGACCGTCCTCGACAAACTCACCTACGCGGCAGCCTCGCCCGCCGCAC 3478
          ||||  |   ||| |  |  ||||  |||||  |    |||||  |||||
Db   1909 GTGACGCAGGTTACGTCGTCGCACAAACTGGGTTACGGAGGCAATCTCAGAAATCTCGCC 1850

Qy   3479 GCGGTGCGTGACCGTCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTC 3538
          |  |  ||   ||   || ||||  ||||| | ||| |||||||||| ||| |||
Db   1849 GAAGCGCTCGGCGGGACCCTCGTCAGCAGTTCGGGGCGACATCTGTGACCGAAGTCTA 1790

Qy   3539 GTCGACACGCTGGCCCGCGGACGCGACGACATCGTGCACTTCGGCGGCCGAGTCGCACGTC 3598
          ||||     |||||||||||    || | ||| | |||||  | ||||| |||||
Db   1789 ATCGAGGGGCTGATGGGCCGGCACGACACCGTGGCGCACTTCGCCGCCGAGACCCACGTC 1730

Qy   3599 GACCCGCTCCATCACCGACAGCCGTGCCTTCACCCGGCCACCAGTCTGGCCACCCAGGTC 3658
          ||||||||    |  |||||||| | |||||  |||| | ||| |||||
Db   1729 GACCCGCTCGGTGTGCCGTCCGCCACGTGTGCGCACCTTCGTGCGCACTCGGGCACTCAGGTG 1670

Qy   3659 CTGCTCGACGCCGGCTCCGCCCACGTGTGCGACGTCTTCGTGCACGTCTCCACCGACGAG 3718
          | - || |||||  ||||     || | ||||  ||||  ||
Db   1669 CTACTGGACGCGCGCTACGCCGCACCATATCGGCCCGTTCCTGCATGTTCCACCGACGAG 1610

Qy   3719 GTGTACGGCTCCCTCCCCCGACAGGGCCGGCCGCGGGAGCGCAGCCCCCTGCTTCCGACCTCG 3778
          |||||       ||  | |    ||||  |||||||  |||  |||||| |||| ||
Db   1609 GTGTACGGGTCGATCGAACCGGCTCGTGGCCGAGGGCCATCCGGCCGCCCCAACTCG 1550

Qy   3779 CCGTACGCCGGCGTCGAAGGCGCGCCTCGACCTCATGGCGGCCACCGCACCCAC 3838
          ||||||  ||| | | || |||  |  || | ||||  | ||||
Db   1549 CCGTACGCCGAGCAAGCCGGGTCCGACCTCTCGCTCTGCCTCTACCACCAGACGCAC 1490

Qy   3839 GGCCTGGACGTCCGGGTGACCCGCTGTTCGAACAACTTCGGCCCCCACCCAGATCCCGAG 3898
          |||||||  | ||||  |||||  | ||||  |  |||  | ||||||
Db   1489 GGGATGGACGTCGGTGGTGACCGGCTCCAACAACTACGGGCCCCGCAATTCCCGGAG 1430

Qy   3899 AAGCTCATACCGGCCGCTTCCTGACCAGCC TCCTGTCCGGCCACCGTTCCCCTCTACGGC 3958
          | - || ||  |  ||  || || - | ||| | || |||||||
Db   1429 AAAATGATTCCGGTTGTTCGTCACCAGGCTGCTGCGACGGGCTCGACGGGCTCTACGGC 1370

Qy   3959 GACGGGCGGCACGTGCCCGACTGGCTGCGACGTCGACGACCACGTCAGGGCCGTCGAACTC 4018
          |||||||    ||| | | |  ||||||  |   |||| | |  |||  ||| |  |||
Db   1369 GACGGCCGCAACATCCGGCCTGGCTCCACGTCCACGTCAGCGACCATTGCCCGCCCTG 1310

Qy   4019 GTCCGCGTGTTCGGGCGCCGGCGGCGGGAGAGATCTACAACATCGGGCGGCACCTGCTGCCC 4078
          | - || |||| |  ||||||||||||||     || |   |  |  ||
Db   1309 GCCCTGGTGTCCGCCGGGCCGGCGGAGGCCAGGTCTTATCATCGGCGGGTGGGAGGCGACG 1250

Qy   4079 AACCTGGAGCTCACCGACGTCGCCCGGTTGCTCCGCACCGTGTCGGCCGCGGGAGCCGCATCGTC 4138
          | |||| |  ||  ||||  |||||||    |||||
Db   1249 AATCTCGAATTGACCGAGATCCTCCGCGAGGCGTGGGCGCCCCGGTCGTTGGGACTCAGATATCT 1190

Qy   4139 CACGTCGAGACCGCAAGCAGCAGACCGCAAGTCACGACCGGCTCATTCTCTGACTAATTCGAAAATCGCC 4198
          |  | |||    ||  |||   | | ||||| | |  |||| |  |  |||
Db   1189 TTCGACGATCGCAAAGTCACGACCGCCGGGTCGGGGCGCTATCCTCGTCGACTAATTCGAAAATCGCC 1130
```

RESULT 12
US-09-194-905-4/c

; Sequence 4, Application US/0919405
; Patent No. 6306627
; GENERAL INFORMATION:
;  APPLICANT: DECKER, Heinrich
;  TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
;  TITLE OF INVENTION: GLA.O AND THEIR USE
;  NUMBER OF SEQUENCES:   13
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FOLEY & LARDNER
;   STREET:  3000 K Street, N.W.
;   CITY:  Washington
;   STATE:  D.C.
;   COUNTRY:  U.S.A.
;   ZIP:  20007-5109
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/194,905
;   FILING DATE: 29-JUL-1998
;   CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/EP97/02826
;   FILING DATE: 30-MAY-1997
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: DE 19622783.6
;   FILING DATE: 07-JUN-1996
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Granados, Patricia D.
;   REGISTRATION NUMBER: 33,683
;   REFERENCE/DOCKET NUMBER: 026083/0193
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (202) 672-5300
;   TELEFAX:  (202) 672-5399
; INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH:  541 base pairs
;   TYPE:  nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY:  linear
;  MOLECULE TYPE: DNA (genomic)
US-09-194-905-4

Query Match         4.0%;    Score 276.2;  DB 4;   Length 541;
Best Local Similarity 69.6%;  Pred. No. 4.7e-32;
Matches  374;  Conservative   0;  Mismatches 163;  Indels   0;  Gaps   0;

```
Qy   3362 GCGGGCTTCATTCCTCCGCTACGTCCCGCTCCTGTCGCCCCGGCGCCCCGGCCGGGC 3421
          ||||| |||| ||||   ||| | ||||  |||||||
Db    537 GCGGGGTTCATGGGCTCGCACTAGTCCGCACCCTCCTCGGTCCGGACGGCCCCGAC 478

Qy   3422 GTCGCGGTGACCGTCCTGACCAAACTCACCTACGCGGCAGCCTCGCCCGCTCACGCGG 3481
          | |   |||  | |||||||||   ||||||  | |||| || || ||||
Db    477 GCCGTCGTCACCGTGCTCGACGCCCTGAGCTACCGGCTACGCCGGCCAACCTCGCCACCTCGACCCG 418

Qy   3482 GTGCGTGACCATCCCGGCCTTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTCGTC 3541
```

```
Qy  3576  ACTTCGCGGCGAGTCGACGTCGCACGTCGACGGTCCATCACCGACGACGGCGGTCCTTCACCCGCA  3635
          ||  || |||  ||  ||| ||   ||||||| ||  ||||| ||||||||||||||||||||
Db  3867953 ATTTTGCCGCGCAATCCCATGTCGCAATGACACTGACAATGACACAATCCGGAGCCGTTTCTGCACA  3868012

Qy  3636  CCAACGTGCTGGGCACCCAGGTCCTGCTCACGCGCTCCGCGCCGTGTGCGCACCT  3695
          |||||||  | |||||| |||||| |  ||  ||||||||||||||||||
Db  3868013 CCAACGTCATCGGGACCTTCACCATCTGGAACGGTGCGACGCCCACGGGTGCGCG---CC  3868069

Qy  3696  TCGTGCACGTCTCCACCGACGAGGTGTACGGTCCCTCCCGCGCCGCGCGGAGA  3755
          |||||| |  ||||||  ||||| ||||||||  ||||||||||||| |
Db  3868070 TGCACCACATCTCCACCGACGAGGTCTCACGCGCACTTGGAGCTCGACGCGGCGGGT  3868129

Qy  3756  GCGGAC------CCCTGCTTCCGGACCTCCGACCTACGCGCCGTCGAAGGCGGCCTCGG  3806
          |  ||       ||  ||||||||||||||   ||||| |  | || |||||
Db  3868130 TCACCGAATCACGCCCTATAACCGTCCAGCCCCTACTCGGCGCCTGCGACGTCGCCGACCGCCAG  3868189

Qy  3807  ACCTCATGGCGCTCGCCCACCGCCACCGCGCTCGCCCCAACAGCCCC  3866
          ||||  |  ||||||| || || ||| |||  ||||| | ||||||||
Db  3868190 ACATGTTGGTCCGGGCTGGGTTCGGTCCTATGGCGCCTATGGCGGACGATCTCCAACTGCT  3868249

Qy  3867  CGAACAACTTCGGCCCCCACCAGCATCCCACCGACGAGTTCATACCGGCGCTTCCTGACCAGCC  3926
          |||||||  ||||| ||| || |  | |||||||| ||  ||||||||| ||||||||||||
Db  3868250 CCAACAACTACGGCCCGCTATACAGCGCGGGCGTATCAGCACGTCGAGAAGTTCATTCCGCGTCAGATCACCAATG  3868309

Qy  3927  TCCTGTCCGCGGCCGCGCCGCGCGCCGGGGCCCAAGCTCTACGGCGCGCGGCGCCCAATG  3986
          |||| || | |||| ||||| || |  ||||||||||||||||||||  ||||| |  |||
Db  3868310 TGCTCACCGGGCGGCGGCGGCGGCGGCCCAATGTCCGTGACTGACTGC  3868369

Qy  3987  AGTGTGACGACCACGTCAGGGCGTCAGGGCCGTCGAACTCGTCCGCGCTGTGGGCGCGGAGAGA  4046
          ||||||||| |  ||| |||||||  |||  || || ||||||  || |||||||||
Db  3868370 ACGTCGACGACCACACAGCCGCGGTGCCGGAATCCTGGACAGAGGCCGCATCGGCCGAA  3868429

Qy  4047  TCTACAACATCGGGGCGCACCCGCTGCCCAACCGTGCGAGTCCACGCACCGGTTGCTCG  4106
          |    ||||||    ||| ||| |   ||||||| |||  ||  ||||| | | |||
Db  3868430 CCTACCTGATCAGCTCGAGGCGAGCGTGACAACCTGACCCGTCTCGCACCGTGCTGC  3868489

Qy  4107  CACTGTGGCGGCGGCCCGGAGCGTCGAGCTCCACGTCGAGAACCGCAAGGGCACGACC  4166
          || ||||| || || || |||||||||| ||||||||| |   ||||||| ||  |||
Db  3868490 GACTGATGGACCGCGATCCGACGACTTCGACCACGTCACCCGACCGCTCGGCCGCCACGACC  3868549

Qy  4167  GGGGCTACGCGGTCGACCAGCAGAGATCACCCGGAACTCGTTACCGGCC---GCCCA  4223
          |||||||| ||  || |||   |     |||||| ||||||| |||||
Db  3868550 TGCGCTATGCCATCGACCCGTCCACGCTCTACGACGAATTATGCTGGGCGCCAAAGCATA  3868609

Qy  4224  CCGACTTCGCGACCGCCTGGCCGACACCGCGGAAGTGTACGAGAGCCGCACGAGGACTGGT  4283
          |||| | | ||| || ||| |   ||   |||   |||||||  || | |   |||||||||
Db  3868610 CCGATTTCGAGAGAGGCCTGCCGGACACTGACTGGTACCCGCAACGAATCGTTGGT  3868669

Qy  4284  GGCGTCCCTGCTTCCGCCGCGAC  4305
          ||||||  |||||||||| ||||
Db  3868670 GGCGTCCACTAAAGACGCCAC  3868691
```

```
Db  417   GTCCCGACCACCCCCGGCTGCGCTTCGTCGTCGCACGGCGACATCTGCGGCCGACCTGGTC  358
          ||  | ||||||  ||| ||   |||||||||  ||||||||||  ||||  ||||| |||
Qy  3542  GACACGCTGGCCGCGGCGGCGACAGACCATCGTGCCACCGGCACTTGCGGCCGGACTCGCACGTCGAC  3601

Db  357   GACCGGGTCATGGCCGGCCAGGTCGTGCGTGCCACCAGCAGTGCGTCGCACGTCGCACGTGGAC  298
          ||||  | |||||||  ||| ||| ||| ||  ||  || |||| |||||||||||
Qy  3602  CGCTCGCTGCTCACGACGAGCCGGCCGCGCGCTTCCACCAGCTGTCGGCCACCGCCACCGACACTC  3661

Db  297   CGCTCGCTGCTCACCGACGGCGCCGCGCTTCGTCGGACCAACGCGGCCAACACCCGGCCACGACACTC  238
          |||||||| | |||||| |  | |||||||||||||||  |||  ||| |||||||||
Qy  3662  CTCGACGCGCGCTCCGCCACGTGTGCGCACTTCGTGCGACGTCTCCACCGACGAGGTG  3721

Db  237   CTCGACGCGGCGCCCTGCGGACGCGGCGTGGCCCCTTCGTCGTCGCACGTCCACCGACGACACGGTC  178
          ||||| |||  |||||  | | | ||   |  |||||  || |||||| |||  | |||
Qy  3722  TACGGCTCCCCTCCGCCACCGGGGCCCGCGGAGAGCACCCCCTGCTTCCGACCTCCGACCTCGCCG  3781

Db  177   TACGGCTCGTGGACGGGTCGTGACCGAGCCGCCGGCTCGCCCACCACCAAAGCCCC  118
          ||||||| | ||| || || |   ||| || | |||| ||||||||||||
Qy  3782  TACGGCGGCGTCGAAGGCGGCGCCTCATGGGCCTCGCGCTCCCCACCGCACCCACCAGCC  3841

Db  117   TATGCCACCTCCAAGGCCGTCGGGGCACCGTCGCTGCCCGCTGCTCCGCCCACGTCAGCCACCAGGC  58
          ||||||||||||||| ||   ||| | |  |  | ||||||||| ||  |||||||||
Qy  3842  CTGGACGTCCGGGTGACCCGTGTTCGAACAACTTCGGCCCCCACCAGCATCCCGAG  3898

Db  57    CTGGACGTGCGCATCACCCGCTGCTCCAACAACTACGGCCCCCTACCACCCGGGG  1
          |||||||||||| ||||||||||  || |||||||||||||  | |||
Qy  3336  ...
```

Query Match    3.5%;    Score 238;   DB 4;   Length 4403765;
Best Local Similarity   56.5%;    Pred. No. 1.4e-26;
Matches 555;   Conservative 0;   Mismatches 400;   Indels 27;   Gaps 5;

```
Qy  3336  CCACGACCATCCTCTGCTCACCGGCCGGAGCCGGCGGCTTCATTCGCTCCGCTCACGTCCCGCCGGC  3395
          || ||| |  ||| ||| || || |||  ||| ||||||||   || ||  |
Db  3867725 CCATGCGGTTGCTAGTCACCGGTTGGCGCGGGGGATTCATCGGCACGCACCGAATTTCGTGCACAGCG  3867784

Qy  3396  TCCTGTGCTCGGCCGGGGCCCCGGGCGGTCCGGGCCGTCCGGGTGACCGTCGGGCTCCGGTGACCTACG  3455
          |||||| || || || ||||||||||||| ||   |  |||| || ||  ||||||||||||
Db  3867785 CCGT------ACGTGAGCATCCAGACGACGATCGGTTACCGTAGCTCGACCCTGACCTACG  3867838

Qy  3456  CCGGCAGCCTCGCCCCGCTGCACCGGTCGTGCTGCGTGACCATCCCGGCCTCACCTTCGTTCCAGG  3515
          ||||  |||  ||  |  | ||| || ||||| ||   ||||| |||||||||||  ||||| ||||
Db  3867839 CCGGCCCGCGCGAGTCGCTGGCCCACGTCGGAGAT------GCCATCCGCGCTGGTTCAGG  3867892

Qy  3516  GCGACGTGTGCGACCACCGCGCCGCTCGTCGCCGACGACGACATCGTGC  3575
          |||| ||||  ||  |  |  || | | |||||||||| || || ||||
Db  3867893 GCGATATCACCGACGCCGACGCGCCGGTGGTTCCAGCTGGTTGGTCCGAGTCCGGACGACATCGTGGTGT  3867952
```

ZIP: 10016-2391
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/592,874
  FILING DATE:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/377,440
  FILING DATE: 24-JAN-1995
ATTORNEY/AGENT INFORMATION:
  NAME: GOLDBERG, JULES E.
  REGISTRATION NUMBER: 34,408
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 212-986-4090
  TELEFAX: 212-818-9479
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 28804 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: unknown
  TOPOLOGY: unknown
  MOLECULE TYPE: DNA (genomic)
  FRAGMENT TYPE: N-terminal
US-08-592-874-1

```
Query Match         3.4%;     Score 234.8;  DB 2;  Length 28804;
Best Local Similarity 56.4%;  Conservative 0;  Mismatches 377;  Indels 78;  Gaps 5;
Matches 589;

Qy  3330 CCATGACCACGACCATCCTCGTCACCCGGCGGAGCGGGCTTCATTCGCTCCGCCTACGTCC 3389
Db 24681 CCATGCAGCAGACCTTCCTCGTTACCCGGGGCGCCCCCCGGGCGTCGGCGCAGTCGTAC 24740

Qy  3390 GCCGGCTCCTGTCGCCCGGGGCCCCCGGGCGTCGCGGTGCCCGCCGC 3449
Db 24741 GCCACCTCGTTCGTTCGCCAGGGCGCGCGCGGCGTC----------ATCAATCTCGACAAGCTCA 24788

Qy  3450 CCTACGGCCGGCAGCCTCGCCCCGCCTCGCACGGCGGTGCGCCGGGTGCCGCGCGCCTCACCTTCG 3509
Db 24789 CCTATGGCGGGCAACCCGGCCCTGCTCGCCGATCGGAGAACGCCCCCAACTACCGGCTTCG 24848

Qy  3510 TCCAGGGCGACGTGTCGACACCGGCGGCTCGTCGACGACGCTG------GCCGCGCGGCACG 3563
Db 24849 TCCACGCCGATATCGCCGACACCGCGGCTGCGCCGTGCGCGAAGAGCAGGTCG 24908

Qy  3564 ACGACATCGTGCACTTCGCGGCGAGTCGCACGTCGACGCGCTCCATCACCGACAGCGGGTG 3623
Db 24909 AGTGGTGATGCACCTCGCGCCGAGAGCCATGTCGATCGCTCGATCGCGCGGCCGCGCCGGGCG 24968

Qy  3624 CCTTCACCCGCCACCAAGCTGTCTGGGCACCCAGGTCTTCGCTCGCTGCGAGCGCCGCGCTCCGCCA-- 3681
Db 24969 AGTTCATCGAGACCAACGTCGTCGGCCACCTTCAAGCTGCTCCAGGCGGCCGCTGCAATATT 25028

Qy  3682 ------------------------CGGTGTGCGCACCTTCGTGCACGTCGCACGTCTCCACCGACG 3716
Db 25029 GGCGCGAGCTGAAGGGAGAAGCGCGAGGCGTTCCGCTTCCACCACCATTCCACCGACG 25088

Qy  3717 AGTGTACGGCTCCCTCCGCACGGCGCGCGGAGAGCGA------CCCCTGCTTC 3770
Db 25089 AGTGTGTTCGGCGACCTGCCGTTCGACACGCGGCATCTTCACCGAAGAGACGCGCCCTATGATC 25148

Qy  3771 CGACCTCGCGCGTACGGCGGCGTCGAAGGCCGCCTCGAAGGCCCATGGGCGCCTCGCCGCCCACCACC 3830
Db 25149 CCTCCTCGCCCATTCGGCGTCGAAGGCGTCGAAGGCGGCCGCCGCTGGGTCGCGGCCGGGTC 25208

Qy  3831 GCACCCACGACCTGGACGTCCGGTGACCGCGTGTTCGAACAACTTCGGCCCCAGC 3890
Db 25209 ACACCTATGGCGGCAACCCGGCCTCGCCGGATCGGAGACGACGCCCCCAACTACCGGCTTCG 25268
```

```
Qy  3891 ATCCCGAGAAGCTCATACCGGCTTCCTGACCAGCCTCCTGTCCGGCGGCACCGTTCCCC 3950
Db 25269 TCCCGAGAAGCTGATCCCGCTGTGACCATCCTCAACGCGCTGGAAGCAAGCCCCTGCCCG 25328

Qy  3951 TCTACGGCGACGGGCGGCGACTGCGGACTGCCGCACGTGCAGTCGACGACGACCAGTCAGGGCCG 4010
Db 25329 TCTACGGCCAAGGGCGGAGAATATCCGCGACTGCGTCGACGATCACGGCCAAGGCGC 25388

Qy  4011 TCGAACTCGTCCGCGTGTCGGGCGGCCGCCGGGGAGAGATCTACACATCGGGGGGCCACCT 4070
Db 25389 TGGCGACGATCCGACCGGCCAAGGTCGGCCAGAGCTACATGTCGGCGGCCGCCAACG 25448

Qy  4071 CGCTGCCCAACCTGGAGCTCACGCACCGGTGCTCCGCACTG------------- 4111
Db 25449 AGCGCACCAAACCTGCAGTCGTGAGACGATCTCGCTCGCTGCGATCAGCAGCCGCATTCCGC 25508

Qy  4112 -------TGCGGCGCGGGCCCGGGCCCCCATCGTCCACGTCGAGAGACCGAGAGAACCAGG 4163
Db 25509 TGAAGGATGGCAAGAAGCGCCGGAGCTGATCACCTTCGTCACCGATCGCCCCGGCCATG 25568

Qy  4164 ACCGGCGTACGGGTGCGACCAGCAAGATCACCGGGAACTCGGTTACCGGCGCGCCA 4223
Db 25569 ACCGGCGCTACGCGATCGAGCGCGGCCAACCAGTCGAACTGGGCTGGGCTGGAAGGCCGAGG 25628

Qy  4224 CCGACTTCGCGACGCCGCTGGCCGACACCGCGAAGTGTACGAGCGGCGACGAGGACTGGT 4283
Db 25629 AGAATTTCGACACCGGCATCCGCGGCGACGATCCGACTCGTATCTCGAGAATGAATGGTGT 25688

Qy  4284 GGCGTCCCTGCTCCGCGCGACAT 4307
Db 25689 GGGTCCGATCCGCCTCCGGCAAAT 25712
```

```
Query Match         3.4%;     Score 234.8;  DB 3;  Length 28804;
Best Local Similarity 56.4%;  Pred. No. 4.1e-26;
Matches 589;  Conservative 0;  Mismatches 377;  Indels 78;  Gaps 5;

Qy  3330 CCATGACCACGACCATCCTCGTCACCCGGCGGAGCGGGCTTCATTCGCTCCGCCTACGTCC 3389
Db 24681 CCATGCAGCAGACCTTCCTCGTTACCCGGGGCGCCCCCCGGGCGTCGGCGCAGTCGGTAC 24740

Qy  3390 GCCCGGCTCCTGTCGCCCGGGGCCCCCGGGCGTCGCGGTGCCCGCCGC 3449
Db 24741 GCCACCTCGTTCGTTCGCCAGGGCGCGCGCGGCGTC----------ATCAATCTCGACAAGCTCA 24788

Qy  3450 CCTACGGCCGGCAGCCTCGCCCCGCCTCGCACGGCGGTGCGCCGGGTGCCGCGCGCCTCACCTTCG 3509
Db 24789 CCTATGGCGGGCAACCCGGCCCTGCTCGCCGATCGGAGAACGCCCCCAACTACCGGCTTCG 24848

Qy  3510 TCCAGGGCGACGTGTCGACACCGGCGGCTCGTCGACGACGCTG------GCCGCGCGGCACG 3563
Db 24849 TCCACGCCGATATCGCCGACACCGCGGCTGCGCCGTGCGCGAAGAGCAGGTCG 24908

Qy  3564 ACGACATCGTGCACTTCGCGGCGAGTCGCACGTCGACGCGCTCCATCACCGACAGCGGGTG 3623
Db 24909 AGTGGTGATGCACCTCGCGCCGAGAGCCATGTCGATCGCTCGATCGCGCGGCCGCGCCGGGCG 24968

Qy  3624 CCTTCACCCGCCACCAAGCTGTCTGGGCACCCAGGTCTTCGCTCGCTGCGAGCGCCGCGCTCCGCCA-- 3681
Db 24969 AGTTCATCGAGACCAACGTCGTCGGCCACCTTCAAGCTGCTCCAGGCGGCCGCTGCAATATT 25028

Qy  3682 ------------------------CGGTGTGCGCACCTTCGTGCACGTCGCACGTCTCCACCGACG 3716
Db 25029 GGCGCGAGCTGAAGGGAGAAGCGCGAGGCGTTCCGCTTCCACCACCATTCCACCGACG 25088

Qy  3717 AGTGTACGGCTCCCTCCGCACGGCGCGCGGAGAGCGA------CCCCTGCTTC 3770
Db 25089 AGTGTGTTCGGCGACCTGCCGTTCGACACGCGGCATCTTCACCGAAGAGACGCGCCCTATGATC 25148

Qy  3771 CGACCTCGCGCGTACGGCGGCGTCGAAGGCCGCCTCGAAGGCCCATGGGCGCCTCGCCGCCCACCACC 3830
Db 25149 CCTCCTCGCCCATTCGGCGTCGAAGGCGTCGAAGGCGGCCGCCGCTGGGTCGCGGCCGGGTC 25208

Qy  3831 GCACCCACGACCTGGACGTCCGGTGACCGCGTGTTCGAACAACTTCGGCCCCAGC 3890
Db 25209 ACACCTATGGCGGCAACCCGGCCCTGCTCGCCGATCGGAGACGACGCCCCCAACTACCGGCTTCG 25268
```

```
Qy   3510  TCCAGGGGCACGTGTGGACACCGCCGCTCGTCGTCGACACGCTG------GCCGCGCGGCACG  3563
           |||||-|||-|-|        |||||||||||        ||-|||        |||-||-||
Db  24849  TCCACGCCGATATCGCCGACACCGCGACGATCGTCCGCGTGCTGCCGGAAGAGCAGGTCG  24908

Qy   3564  ACGACATCGTGCACTTCGCGGCGCGAGTCGCACGTCGACCGCTCCATCACCGACAGCGGTG  3623
           |||-| ||||||||-||||| ||||-| |||||| |||-||-|||        ||||||||-|
Db  24909  ACGTGGTGATGCACCTCGCCGCCGAGAGCCATGTCGATCGCTCGATCGACGGCCCGGGCG  24968

Qy   3624  CCTTCACCCGCACCACCAACGTGCTGGGCACCCAGGTCCTCGACGTCGACGCCGCTCCGCCA--  3681
           |||||-||||||| -|||||| |-|||-| | -||||        ||||| |-||||
Db  24969  AGTTCATCGAGACCAACGTCGTCGGCACTCGCGGCCACCTTCAAGCTGCTCCAGGCGGCCGCTGCAATATT  25028

Qy   3682  ------------CGGTGTGCGGCACCCTCGTCGCACGTCTCCACCGACG  3716
                    ||||        ||-|||-|||||||||||
Db  25029  GGGCGGAGCTGGAAGGGAGAAGCGCGAGCTTTCCGCTTCCACCACATTTCCACCGACG  25088

Qy   3717  AGGTGTACGGCTCCCTCCCCGCACGGGGGCGCGGAGAGCGA------CCCCCTGCTTC  3770
           |||||| |||| ||| |||-||| ||| |||        |||
Db  25089  AGGTGTTCGGCGACCTGCGTTCGACAGCGGCATCTTCACCGAAGAGACGCCCTATGATC  25148

Qy   3771  CGACCTCGCCGTACGGCGGGTCGAAGGCGGCCTCGGACCTCATGGCGCTCGCCCACCACC  3830
           |-||||||  |||||||||||        ||||-| -||||
Db  25149  CCTCCTGCCCTATTCGGCGTCGAAGGCGGCCAGCGACCATCTGTTCCGCGGCCTCGGGGTC  25208

Qy   3831  GCACCCACGGCCTGAGCTGCGGCTCCGGTGACCCGGCTGTTCGAACAACTTCGGCCCCCACCAGC  3890
           |||| -|-||||||        |-|||        ||-|||||||||||| -||-  |||
Db  25209  ACACCTATGGCCTGCCCGTGGTCGTGTGTCCGAACTGCTCGCGAACAATTACGGCCGTTCCACT  25268

Qy   3891  ATCCCGAGAAGCTCATACCGCGCTTCCTGACCAGCCTCCGGGCGCACCGTTCCCC  3950
           |||||||||||-| |||| ||-||||-|||        ||| |-|||||||||||-|
Db  25269  TCCCCGAGAGCTGCGATCGACGCGCTGGAAGGCAAGCCCCTGCCCG  25328

Qy   3951  TCTACGGCCACGGCGGCGGCACGTCGGCGGACTGGCTCACGTGCAGGACCACGTCAGGGCCG  4010
           |||||| |-|||-|-|||-| ||||-||        ||||||||| -||||| ||-||
Db  25329  TCTACGGCAAGGGCGGAGAATATCCGCGACTGGCTGTACGTCGACGGATCACGCCAAGGCGC  25388

Qy   4011  TCGAACTCGTCCGCGTGTCGGGCGGCCGGAGATCTACAACATCGGGGGCGGCCACCT  4070
           -|-| -|-|||| -|||||| |||| |||        ||-|||||| |||-||||-|
Db  25389  TGGGCGACGATCGGCCACGACCGGCAAGGTCGGCCCAGAGCTACAATGTCGGCGGCCCGCAACG  25448

Qy   4071  CGCTGCCCAACCTGGAGCTCACGCACCGGTTGCTTCCCACTG-----------  4111
           |-| ||||||||||| ||-|       |-| -|-|||
Db  25449  AGCGCACCACCCTGCAGGTCGTCGAGACGATCTGCGACCTGCTCGATCAGCGCATTCCGC  25508

Qy   4112  -------TGCGGCGCGGGCCCGGCATCGTCACCGTCGAGAACCCGCAAGGGGCACG  4163
                  ||-|||| |||||-|||        |||-| ||| -|||
Db  25509  TGAAGGATGGCAAGAAGCGCCCGCACCGGCGCTGATCACCGGTTACCGGCCGCGCA  25568

Qy   4164  ACCGGCGCTACGGCGGTCGACCAGCAAGATCACCCGGGAACTCGGTTACCGGCCGCGCA  4223
           |||||||||||||||-|||        -|||| -||||        ||| |||| -|||||-|
Db  25569  ACCGGCGCTACGGCGATCGACGCGACCAAGCTCGAGACCGAGACTCGGGCTCGGAAGGCCGAGG  25628

Qy   4224  CCGACTTCGCCGACCGCGCTGGCCGACACCGGAACTGGTACGAGCGGCCACGAGGACTGGT  4283
           |||| -|||| ||||||| -| -| ||-|||        ||||        ||-|||
Db  25629  AGAATTTCGACACCGGCATCGCCGGCGATCGACTGGTATCTCGAGAATGAATGGTGGT  25688

Qy   4284  GGGGTCCCCTGCTCGCCGGCGACAT  4307
           ||-||||        |||-|||  |-||
Db  25689  GGGGTCCGATCCGGTCCGGCAAAT  25712
```

Search completed: May 29, 2003, 06:35:06
Job time : 34283 secs

```
              GenCore version 5.1.4_p5_4578
         Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:  May 28, 2003, 19:47:36 ; Search time 500 Seconds
                                  (without alignments)
                               18100.897 Million cell updates/sec

Title:           US-09-922-683-7
Perfect score:   6854
Sequence:        1 CTGCAGGGTTCCCTGGTGCA.........TCGAGCCCGATCCGGCTGCAG 6854

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters:   1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Published_Applications_NA:*
      1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
      2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
      3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
      4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
      5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
      6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
      7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
      8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
      9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
     10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
     11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
     12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
     13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
     14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

## SUMMARIES

| | Result No. | Score | % Score Match | Query/DB Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| | 1 | 6854 | 100.0 | 6854 | 9 | US-09-922-683-7 | Sequence 7, Appl1 |
| c | 2 | 527.4 | 7.7 | 12441 | 9 | US-09-988-384B-3 | Sequence 3, Appl1 |
| c | 3 | 527.4 | 7.7 | 13613 | 9 | US-09-860-846-3 | Sequence 3, Appl1 |
| c | 4 | 527.4 | 7.7 | 13613 | 9 | US-09-836-821-3 | Sequence 3, Appl1 |
| c | 5 | 527.4 | 7.7 | 13613 | 10 | US-09-861-289-3 | Sequence 3, Appl1 |
| c | 6 | 522.6 | 7.6 | 546 | 9 | US-09-922-683-3 | Sequence 3, Appl1 |
| | 7 | 447 | 6.5 | 1014 | 9 | US-09-860-846-13 | Sequence 13, Appl |
| | 8 | 447 | 6.5 | 1014 | 9 | US-09-988-384B-13 | Sequence 13, Appl |
| | 9 | 447 | 6.5 | 1014 | 9 | US-09-836-821-13 | Sequence 13, Appl |
| c | 10 | 447 | 6.5 | 1014 | 10 | US-09-861-289-13 | Sequence 13, Appl |
| c | 11 | 276.2 | 4.0 | 541 | 9 | US-09-922-683-4 | Sequence 4, Appl1 |
| | 12 | 205.2 | 3.0 | 1131 | 9 | US-09-738-628-373 | Sequence 373, App |
| c | 13 | 163.6 | 2.4 | 1344 | 9 | US-09-738-626-536 | Sequence 536, App |
| | 14 | 157 | 2.3 | 879 | 9 | US-09-860-846-11 | Sequence 11, Appl |
| | 15 | 157 | 2.3 | 879 | 9 | US-09-988-384B-11 | Sequence 11, Appl |
| | 16 | 157 | 2.3 | 879 | 10 | US-09-836-821-11 | Sequence 11, Appl |
| | 17 | 157 | 2.3 | 879 | 9 | US-09-836-821-11 | Sequence 11, Appl |
| c | 18 | 155 | 2.3 | 1383 | 10 | US-09-815-242-4002 | Sequence 4002, Ap |
| | 19 | 131.4 | 1.9 | 536165 | 9 | US-09-939-964-1 | Sequence 1, Appl1 |
| | 20 | 131 | 1.9 | 855 | 9 | US-09-738-626-371 | Sequence 371, App |
| c | 21 | 130 | 1.9 | 1338 | 10 | US-09-815-242-7604 | Sequence 7604, Ap |
| c | 22 | 125.4 | 1.8 | 4257 | 9 | US-09-825-288A-1 | Sequence 1, Appl1 |
| | 23 | 122 | 1.8 | 15872 | 9 | US-09-860-846-1 | Sequence 1, Appl1 |
| | 24 | 122 | 1.8 | 15872 | 9 | US-09-988-384B-1 | Sequence 1, Appl1 |
| | 25 | 122 | 1.8 | 15872 | 9 | US-09-836-821-1 | Sequence 1, Appl1 |
| | 26 | 122 | 1.8 | 15872 | 10 | US-09-861-289-1 | Sequence 1, Appl1 |
| c | 27 | 122 | 1.8 | 88421 | 9 | US-09-976-059-1 | Sequence 1, Appl1 |
| | 28 | 120.2 | 1.8 | 1221 | 10 | US-09-815-242-7711 | Sequence 7711, Ap |
| | 29 | 118.6 | 1.7 | 13842 | 9 | US-09-860-846-30 | Sequence 30, Appl |
| | 30 | 118.6 | 1.7 | 13842 | 9 | US-09-988-384B-30 | Sequence 30, Appl |
| | 31 | 118.6 | 1.7 | 13842 | 9 | US-09-836-821-30 | Sequence 30, Appl |
| | 32 | 118.6 | 1.7 | 13842 | 10 | US-09-861-289-30 | Sequence 30, Appl |
| | 33 | 118.6 | 1.7 | 36778 | 9 | US-09-860-846-5 | Sequence 5, Appl1 |
| | 34 | 118.6 | 1.7 | 36778 | 9 | US-09-836-821-5 | Sequence 5, Appl1 |
| | 35 | 118.6 | 1.7 | 36778 | 10 | US-09-861-289-5 | Sequence 5, Appl1 |
| | 36 | 118.6 | 1.7 | 37948 | 9 | US-09-988-384B-5 | Sequence 5, Appl1 |
| c | 37 | 118 | 1.7 | 4039 | 10 | US-09-205-448-7 | Sequence 7, Appl1 |
| | 38 | 118 | 1.7 | 11220 | 9 | US-09-860-846-32 | Sequence 32, Appl |
| | 39 | 118 | 1.7 | 11220 | 9 | US-09-988-384B-32 | Sequence 32, Appl |
| | 40 | 118 | 1.7 | 11220 | 9 | US-09-836-821-32 | Sequence 32, Appl |
| | 41 | 118 | 1.7 | 11220 | 10 | US-09-861-289-32 | Sequence 32, Appl |
| c | 42 | 117.6 | 1.7 | 927 | 10 | US-09-815-242-4095 | Sequence 4095, Ap |
| c | 43 | 117.2 | 1.7 | 88421 | 9 | US-09-976-059-1 | Sequence 1, Appl1 |
| c | 44 | 116.8 | 1.7 | 536165 | 9 | US-09-939-964-1 | Sequence 1, Appl1 |
| | 45 | 116.6 | 1.7 | 38734 | 9 | US-09-373-658-30 | Sequence 30, Appl |

## ALIGNMENTS

```
RESULT 1
US-09-922-683-7

; Sequence 7, Application US/09922683
; Publication No. US20020192793A1

GENERAL INFORMATION:
   APPLICANT: DECKER, Heinrich
   TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
                       PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
                       GLA.O AND THEIR USE
   NUMBER OF SEQUENCES: 13
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: FOLEY & LARDNER
      STREET: 3000 K Street, N.W.
      CITY: Washington
      STATE: D.C.
      COUNTRY: U.S.A.
      ZIP: 20007-5109
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/922,683
      FILING DATE: 07-Aug-2001
      CLASSIFICATION: <Unknown>
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 09/194,905
      FILING DATE: 1999-12-01
      APPLICATION NUMBER: DE 1962783.6
      FILING DATE: 07-JUN-1996
   ATTORNEY/AGENT INFORMATION:
      NAME: Granados, Patricia D.
      REGISTRATION NUMBER: 33,683
      REFERENCE/DOCKET NUMBER: 026083/0193
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202) 672-5300
      TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 7:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 6854 base pairs
      TYPE: nucleic acid
```

```
;
;
;
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-922-683-7

Query Match          100.0%;    Score 6854;   DB 9;   Length 6854;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 6854;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 CTGCAGGGTTCCCTGGTGCACGACCAGCCCCTGGTCGACGACCAGGGCGCTGTCGCAGAT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTGCAGGGTTCCCTGGTGCACGACCAGCCCCTGGTCGACGACCAGGGCGCTGTCGCAGAT 60

Qy    61 CGCGGCGATGTCGGCGATGTCGGCTGGTGAGCACCACGGTGCCCAGTTCCCGGTG 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CGCGGCGATGTCGGCGATGTCGGCTGGTGAGCACCACGGTGCCCAGTTCCCGGTG 120

Qy   121 GGGCGGTTGACCAGCCGGCCACCGCGGTCCTTCAGCACCATGTCGAGGCCGATCGTGGG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGCGGTTGACCAGCCGGCCACCGCGGTCCTTCAGCACCATGTCGAGGCCGATCGTGGG 180

Qy   181 CTCGTCCCAGAACAGCACGCACGCCCGGGTCGTGCAGCAGGCTCGCCGCGATCTCGGCGCGCAT 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTCGTCCCAGAACAGCACGCACGCCCGGGTCGTGCAGCAGGCTCGCCGCGATCTCGGCGCGCAT 240

Qy   241 CGGCTGTCCGAGCTGAGCTGCGCCGCCACGGGGTGGACCCCAGCCGTCGTGATGTCGAGGAG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CGGCTGTCCGAGCTGAGCTGCGCCGCCACGGGGTGGACCCCAGCCGTCGTGATGTCGAGGAG 300

Qy   301 GTCCCGGAACAGGGCGAGGTTGCGCCGGTAGACCGGTCCGGGGATGTCGTAGATGCGGCG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTCCCGGAACAGGGCGAGGTTGCGCCGGTAGACCGGTCCGGGGATGTCGTAGATGCGGCG 360

Qy   361 CAGGATGCGGAAGGAGTCGGGTACCGACAGTCCCACCAGAGCTGGCTGCGGCCGAA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAGGATGCGGAAGGAGTCGGGTACCGACAGTCCCACCAGAGCTGGCTGCGGCCGAA 420

Qy   421 GACGACGCCCATCGTGCGGCCGTTGCGGCCGATGTAGGGCGGTCCGGTAGGGCTCCAGCCGGCGAC 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACGACGCCCATCGTGCGGCCGTTGCGGCCGATGTAGGGCGGTCCGGTAGGGCTCCAGCCGGCGAC 480

Qy   481 CGTGCAGCGGCGCGGAGGTGGGGTCATGATGCCGGTCAGGATCTTGATCGTGGTTCGACTT 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CGTGCAGCGGCGCGGAGGTGGGGTCATGATGCCGGTCAGGATCTTGATCGTGGTTCGACTT 540

Qy   541 GCGGGCTCCGTTGGCCGCCGATGTAGGCGGTTCTTCGCGGCCGTATCTCGAAGGAGAC 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCGGGCTCCGTTGGCCGCCGATGTAGGCGGTTCTTCGCGGCCGTATCTCGAAGGAGAC 600

Qy   601 GTCGTCAGCGGCGCCGGAGGGTGGGGTACCCGGGCGTCAGGAGGGTGGAGAGGCTGCCGAG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GTCGTCAGCGGCGCCGGAGGGTGGGGTACCCGGGCGTCAGGAGGGTGGAGAGGCTGCCGAG 660

Qy   661 CAGGGCCGGCTCGCGTTCGGCGGCGGCCGGAACTCCTTGACGAGGTGTTCGGCCAGGATCAC 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CAGGGCCGGCTCGCGTTCGGCGGCGGCCGGAACTCCTTGACGAGGTGTTCGGCCAGGATCAC 720

Qy   721 GCGATCACCCCGCTCGACGGCGGCGTCTCCAGCAGGCGGCCCTCCGAGCAGCGCCCTCG 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GCGATCACCCCGCTCGACGGCGGCGTCTCCAGCAGGCGGCCCTCCGAGCAGCGCCCTCG 780

Qy   781 TCGAGGGTGAACGGCGGTGTAGACGGCCCGGCCCTCGGGGGCGGCCTTCGGGGGCGGCCTTCCAGG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TCGAGGGTGAACGGCGGTGTAGACGGCCCGGCCCTCGGGGGCGGCCTTCGGGGGCGGCCTTCCAGG 840

Qy   841 AGGTCGAGGGCGGTGGTGTGTAGACGGGCGGTCTCGGGGGCCGGTCTCCGGGGCCGGCCCGGCCGACG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGTCGAGGGCGGTGGTGTGTAGACGGGCGGTCTCGGGGGCCGGTCTCCGGGGCCGGCCCGGCCGACG 900

Qy   901 GCGTCGGTGACGAACTCCAGGGGCCCCACACGTCCGAGGCCGTACCTGGCCGGAGCTGG 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GCGTCGGTGACGAACTCCAGGGGCCCCACACGTCCGAGGCCGTACCTGGCCGGAGCTGG 960
```

```
Qy   961 GGGAAGCGGGACTCCAGGGGCGGCGGCAGCCGCCGCTCCTGGATGAGCTGGCGGCGGCGGCGGCGGCCGGACGGCGCACG 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GGGAAGCGGGACTCCAGGGGCGGCGGCAGCCGCCGCTCCTGGATGAGCTGGCGGCGGCGGCGGCGGCCGGACGGCGCACG 1020

Qy  1021 CGGTCGATCAGCCGGTTCGCGTCGACGACCTCAGCCGTGGCGGCGGCGGCGGCGGATCCCC 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CGGTCGATCAGCCGGTTCGCGTCGACGACCTCAGCCGTGGCGGCGGCGGCGGCGGATCCCC 1080

Qy  1081 AGTGGGTTGCTCGCGTACGTCGAGGCGTACGCCGCCGGGGTGGCCGGCCTTCCGCGCTGCGCA 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 AGTGGGTTGCTCGCGTACGTCGAGGCGTACGCCGCCGGGGTGGCCGGCCTTCCGCGCTGCGCA 1140

Qy  1141 GCTTCCGCGCGGTTCCGGCCAGCAGCGGCGGAAGGGAATCCGCGTCCGCGGTGCCCTTGGACAGC 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 GCTTCCGCGCGGTTCCGGCCAGCAGCGGCGGAAGGGAATCCGCGTCCGCGGTGCCCTTGGACAGC 1200

Qy  1201 ATCGCCAGGTCCGGCTCGATGCCGAACAGTTCGCTGGCGGAGGAAGGGCCGGCCGCCCG 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 ATCGCCAGGTCCGGCTCGATGCCGAACAGTTCGCTGGCGGAGGAAGGGCCGGCCGCCCG 1260

Qy  1261 CCGCCGGTGAGGACCTCTCGGCGACGAGCAGCAGCGCCCGTCCCGCGCCAGGCGCCGGCG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCGCCGGTGAGGACCTCTCGGCGACGAGCAGCAGCGCCCGTCCCGCGCCAGGCGCCGGCG 1320

Qy  1321 ATCCGCTCCCAGTAGCCGGGCGGCGGCCACGATGACGGCTGCCGCGCCGGAGGACGGGGTTCG 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 ATCCGCTCCCAGTAGCCGGGCGGCGGCCACGATGACGGCTGCCGCGCCGGAGGACGGGGTTCG 1380

Qy  1381 AAGACCAGGGCCAGACGTTGGGCTTCTCCGGCGGATGTCCGGCGCACGAGGGTCGCGCAC 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 AAGACCAGGGCCAGACGTTGGGCTTCTCCGGCGGATGTCCGGCGCACGAGGGTCGCGCAC 1440

Qy  1441 CGGACGTCGCACGAGGGGTACTCCAGGCCCAGGGACAGCCGGTAGCCAGTAGCCAGTGGGCAGCATCCGG 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CGGACGTCGCACGAGGGGTACTCCAGGCCCAGGGACAGCCGGTAGCCAGTAGCCAGTGGGCAGCATCCGG 1500

Qy  1501 GCCAGCACGCTGTTGCCGCTGAAGGCCTGGTGGCCGTAGGGCTCCAGCCGGCGAC 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GCCAGCACGCTGTTGCCGCTGAAGGCCTGGTGGCCGTAGGGCTCCAGCCGGCGAC 1560

Qy  1561 GCGGCCCATGGTCTTGCCGTGGCGGAAGCCGTTGCGGCCGCAGATCCGGTTGCGGCCCGGC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GCGGCCCATGGTCTTGCCGTGGCGGAAGCCGTTGCGGCCGCAGATCCGGTTGCGGCCCGGC 1620

Qy  1621 GCGGCGGTCGTCGGCCTGACGACCCGCCAGGCGGCCTCGACCACCTCCGCGCCGGGTGGAGAAG 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 GCGGCGGTCGTCGGCCTGACGACCCGCCAGGCGGCCTCGACCACCTCCGCGCCGGGTGGAGAAG 1680

Qy  1681 AAGGGCTAGGTGTCGAGCTGTTCGAGCTGTCGGCGGTCAGGCCGCGCGGCGGCCGGCG 1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 AAGGGCTAGGTGTCGAGCTGTTCGAGCTGTCGGCGGTCAGGCCGCGCGGCGGCCGGCG 1740

Qy  1741 CGGTCCGGGCGGTCTGGCGGCTGTCGTGGACGTTCCACAGGCGGCGGCGGGTGGTGAGTGCC 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 CGGTCCGGGCGGTCTGGCGGCTGTCGTGGACGTTCCACAGGCGGCGGCGGGTGGTGAGTGCC 1800

Qy  1801 TCGACGACCTCCGGGTGCCCGTGCCCGACTGGTGGCGGCCGAAGTCG 1860
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 TCGACGACCTCCGGGTGCCCGTGCCCGACTGGTGGCGGCCGAAGTCG 1860

Qy  1861 AGGTACTCGGTTGCCGTCGGGTCAGAACGGGACCCGCGTCGGGCGGCCTCCGCGTAAGACCCCG 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGGTACTGGTTGCCGTCGGGTCAGAACGGGACCCGCGTCGGGCGGCCTCCGCGTAAGACCCCG 1920

Qy  1921 CGTCCGTGGACGGCTTCCTCGGACGCGCCTTCGGCGGCCTCCGTGCCCGGCCTGCCAGG 1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CGTCCGTGGACGGCTTCCTCGGACGCGCCTTCGGCGGCCTCCGTGCCCGGCCTGCCAGG 1980

Qy  1981 TGCTGTCGTCTGCCGTAAGCCTGTCATCGCTGTCGTCTCGCTCGTCGTCGTCTGCGGCTGACCGGAT 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 TGCTGTCGTCTGCCGTAAGCCTGTCATCGCTGTCGTCTCGCTCGTCGTCGTCTGCGGCTGACCGGAT 2040
```

```
Qy  2041  CGCCGGCGAACTGCGTCGGTGTGGCGCCACACGGTTGGGCGGCGCTCGGCGCGTGAGTCAAAACAC  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  CGCCGGCGAACTGCGTCGGTGTGGCGCCACACGGTTGGGCGGCGCTCGGCGCGTGAGTCAAAACAC  2100

Qy  2101  TTGAACACACACCGCTGCAGAGTTGGCGGGTTGTTCAGAAAGTTGTTGCGAGCGGCCC         2160
          |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db  2101  TTGAACACACACCGCTGCAGAGTTGGCGGGTTGTTTCAGAAAGTTGTTGCGAGCGGCCC         2160

Qy  2161  CGGGCACTCGGTTCAGTCGACGTGCTTACGGCGACCGCTCCACGTTCGAGGAGGGA            2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  CGGGCACTCGGTTCAGTCGACGTGCTTACGGCGACCGCTCCACGTTCGAGGAGGGA            2220

Qy  2221  CCTGTGAGAACAAGCCCGCAGACCGACCCGCTCCGCGAGGCCGAGTGAAGGCCCTG            2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  CCTGTGAGAACAAGCCCGCAGACCGACCCGCTCCGCGAGGCCGAGTGAAGGCCCTG            2280

Qy  2281  TCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACCACCCGCCAAGCAGC         2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  TCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACCACCCGCCAAGCAGC         2340

Qy  2341  TGCTCCCCATCGCCAAACAGCCCGTCTCGCTCTTCTACGGCGTTGAGTCGCTCCCCGCGGCGG     2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TGCTCCCCATCGCCAACAAGCCCGTCTCGCTCTTCTACGGCGTTGAGTCGCTCCCCGCGGCGG     2400

Qy  2401  GTGTCGGGGAGGCCGGCGTCGTGTGGGCGTCGGATCCGCGAACTCACCG                  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  GTGTCGGGGAGGCCGGCGTCGTGTGGGCGTCGGATCCGCGAACTCACCG                  2460

Qy  2461  GCGACGGCACCGCGTTCGGGTTACGCATCACCTACCTCCACAGCCCGCTCGGCGTC           2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GCGACGGCACCGCGTTCGGGTTACGCATCACCTACCTCCACAGCCCGCTCGGCGTC           2520

Qy  2521  TCGGCCACGCGGTGCGCATCGCCGCGGCTTCCTGGGGCGACGACTTCCTGCTGTACC          2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  TCGGCCACGCGGTGCGCATCGCCGCGGCTTCCTGGGGCGACGACTTCCTGCTGTACC          2580

Qy  2581  TGGGGGACAACTACCTGCCCCGCTGCTCACCCCGGTCTGCGCCGACTTCGCCCAATCGGCGCCGATC    2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TGGGGGACAACTACCTGCCCCGCTGCTCACCCCGGTCTGCGCCGACTTCGCCCAATCGGCGCCGATC    2640

Qy  2641  CCCGGGCGGCCGCCCGGTCGCGGACCCGTCGCGGCGG                               2700
          ||||||||||||||||||||||||||||||||||||
Db  2641  CCCGGGCGGCCGCCCGGTCGCGGACCCGTCGCGGCGG                               2700

Qy  2701  AGTCGACGGCGACGGCGGAACGTGCTGCTGCGCTTGGAGGAGAAACCGAGCTCCGGCCAGCT      2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  AGTCGACGGCGACGGCGGAACGTGCTGCTGCGCTTGGAGGAGAAACCGAGCTCCGGCCAGCT      2760

Qy  2761  CGCTCGCCTCATCGGCGTGTAGCCCGTCGACGAGGCGGTCCCGGCCA                    2820
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  CGCTCGCCTCATCGGCGTGTAGCCCGTCGACGAGGCGGTCCCGGCCA                    2820

Qy  2821  TCACCCCCTCGCGCCCGGCCCGGAGCTGGAGATCCGAGGCCGTCAGTGATCGACC            2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  TCACCCCCTCGCGCCCGGCCCGGAGCTGGAGATCCGAGGCCGTCAGTGATCGACC            2880

Qy  2881  GGGGCTGCGCGTACGGCCGAGACCACCGCGCCCTGGCGCGGCACCGGCAGCCGG             2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  GGGGCTGCGCGTACGGCCGAGACCACCGCGCCCTGGCGCGGCACCGGCAGCCGG             2940

Qy  2941  AGGACATGCTGGAGGTCAACCGCACGTCACGTGGACGAGGGCCCATCGAGGGGA             3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  AGGACATGCTGGAGGTCAACCGCACGTCACGTGGACGAGGGCCCATCGAGGGGA             3000

Qy  3001  AGTGCGCAGCGCACGCCACGGTCGCGGTGGCCGAAGGCGGCGATCGTGC                  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  AGTGCGCAGCGCACGCCACGGTCGCGGTGGCCGAAGGCGGCGATCGTGC                  3060

Qy  3061  GGGGTCACACGTCGTGGGCCCGGGTGATGCGCGGTTGTCGGCGGGTCGTCGTCAGCAACTCCA    3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  GGGGTCACACGTCGTGGGCCCGGGTGATGCGCGGTTGTCGGCGGGTCGTCGTCAGCAACTCCA    3120

Qy  3121  GTGTCGGCGGCCGTACACCTCCATCGGCGGAGGACTGCCGGGTCGAGGACACCGCCAGTG       3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  3121  GTGTCGGCCCGTACACCTCCATCGGCGGCGGCGGCCGGCCGGTCGAGGACACCGCCATCGAGT     3180

Qy  3181  ACTCCGTCCTGCTGGCGGCGGCGGCGGCCCGTCGTCGGCGCCGCCCAGGCAGGGGCGGCGTCCCTCA  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  ACTCCGTCCTGCTGGCGGCGGCGGCGGCCCGTCGTCGGCGCCGCCCAGGCAGGGGCGGCGTCCCTCA  3240

Qy  3241  TCGGCGGCGGCGGCGGCCGTCGTCGGCGGCCCCCGTCTCCCCGCAGGCAGGTCACCGACTGGTGA    3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241  TCGGCGGCGGCGGCGGCCGTCGTCGGCGGCCCCCGTCTCCCCGCAGGCAGGTCACCGACTGGTGA    3300

Qy  3301  TCGGGCGACCACAGCAAGGTGTATCTCACCCCATGACACGAACCATCCTCGTCACCGCGCCGG     3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  TCGGGCGACCACAGCAAGGTGTATCTCACCCCATGACACGAACCATCCTCGTCACCGCGCCGG     3360

Qy  3361  AGCGGGCGTTCATTCGCTCGCCGCTACGTCCTGTGCCGCCGCTCGCCGAGCCTCGCCACGC       3420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361  AGCGGGCGTTCATTCGCTCGCCGCTACGTCCTGTGCCGCCGCTCGCCGAGCCTCGCCACGC       3420

Qy  3421  CGTCGGGTGACCGTCCTCGACAAACTCACCTCCAGGGGCGACGTGCGCACCGCGCACCGCGCT      3480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421  CGTCGGGTGACCGTCCTCGACAAACTCACCTCCAGGGGCGACGTGCGCACCGCGCACCGCGCC      3480

Qy  3481  GGTGCGTGACCATCCCGGCCTTCGTCGCGCCGCGGCACGTGCGACTTCGCGCACCGCGCTCGT     3540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  GGTGCGTGACCATCCCGGCCTTCGTCGCGCCGCGGCACGTGCGACTTCGCGCACCGCGCTCGT     3540

Qy  3541  CGACACGCTGCGCGGCGGCGACGACAGCCGCGACACGTCGACTTCGGCGCACTTCGCACGTCGA     3600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  CGACACGCTGCGCGGCGGCGACGACAGCCGCGACACGTCGACTTCGGCGCACTTCGCACGTCGA     3600

Qy  3601  CCGCTCCATCACCGACGCGTGCCTTCACCCCACCAAGCTGCTGGGCACCCAGCTGCTCCT        3660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  CCGCTCCATCACCGACGCGTGCCTTCACCCCACCAAGCTGCTGGGCACCCAGCTGCTCCT        3660

Qy  3661  GCTCGACGTCCGGCTGTGACGCGTCTTCGTGCGCACCGTGCGCACTTCGCACGGACGAGGT      3720
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  3661  GCTCGACGTCCGGCTGTGACGCGTCTTCGTGCGCACCGTGCGCACTTCGCACGGACGAGGT      3720

Qy  3721  GTACGGCTCCCTCCCCGCTTCCTGCTTCCCGACCTTCCGACCTCGCC                    3780
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  GTACGGCTCCCTCCCCGCTTCCTGCTTCCCGACCTTCCGACCTCGCC                    3780

Qy  3781  GTACGGGCGTCGAAGGCGGCGGCCTCATGGCGTGCGACCTGCTGCACGTCAGGGCCGTCGAACTCGT  3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781  GTACGGGCGTCGAAGGCGGCGGCCTCATGGCGTGCGACCTGCTGCACGTCAGGGCCGTCGAACTCGT  3840

Qy  3841  CCTGGACGTCCGGTGACCCGCTGTTCGAACAACTTCGCCCGCGCATCCCGAGAA              3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841  CCTGGACGTCCGGTGACCCGCTGTTCGAACAACTTCGCCCGCGCATCCCGAGAA              3900

Qy  3901  GCTCATACCGGCGCTTCCTGACCAGCCTTCCTGTCCGACCGTCCCCTCTACCGCGA            3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901  GCTCATACCGGCGCTTCCTGACCAGCCTTCCTGTCCGACCGTCCCCTCTACCGCGA            3960

Qy  3961  CGGGGCGGCACGTGCGCGACCACGTCGCACGTCAGGGCCGTCGAACTCGT                 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961  CGGGGCGGCACGTGCGCGACCACGTCGCACGTCAGGGCCGTCGAACTCGT                 4020

Qy  4021  CCGCGTGTCGGGCGCCGACGTCCGGGGAGAGATCTACAACATCGGGGCGCGGCACCTGCCGCCAA    4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021  CCGCGTGTCGGGCGCCGACGTCCGGGGAGAGATCTACAACATCGGGGCGCGGCACCTGCCGCCAA    4080

Qy  4081  CCTGGAGCTCACGCGCACGGCGATCGTGC                                       4140
          |||||||||||||||||||||||||||||
Db  4081  CCTGGAGCTCACGCGCACGGCGATCGTGC                                       4140

Qy  4141  CGTCGAGAACGGCAAGGGCGTCACCGGCGTACGGCGTCAACTCCA                      4200
          ||| ||||||||||||||||||||||||||||||||||||||||||
Db  4141  CGTCGAGAACCGCAAGGGCACCGGCGTACGGCGTCGTCAACTCCA                      4200

Qy  4201  GGAACTCGGTTACCGGCGCGCGACTTCCGGCGGCGCTGGCCGACACCGCGAGTG             4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  4201  GGAACTCGGTTACCGGCCGCCGCCGACTTCGCGACCGCGCTGGCCGCACCGCGAAGTG  4260
                                                                            
Qy  4261  GTACGACGCGGAGGACTGGTGTGCGTCGGCGCGACGACTGCGGGCCGCCCGGTCGGGGCCGG  4320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4261  GTACGAGCGGCACGAGGACTGGTGTGCGTCGGCGCGACGACTGCGGGCCGCCCGGTCGGGGCCGG  4320

Qy  4321  ACCGCAACCACCGCCCGGCCGGCGACACCGCCGGGGCGCCGGGCGGGCCGCCCGGACCACC  4380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4321  ACCGCAACCACCGCCCGGCCGGCGACACCGCCGGGGCGCCGGGCGGGCCGCCCGGACCACC  4380

Qy  4381  CGTCCGTGAGCCGGGGGCGGCCGGCGCGCCCCGGGGCGCCGGCGGTGACCCCGGACCACC  4440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4381  CGTCCGTGAGCCGGGGGCGGCCGGCGCGCCCCGGGGCGCCGGCGGTGACCCCGGACCACC  4440

Qy  4441  AGTTCCGGCATGAAGACGCAATTCGGTTGCGCGACCTGCCCATCGCCTCGATGGTGTCAGGGGA  4500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441  AGTTCCGGCATGAAGACGCAATTCGGTTGCGCGGCCGTTCCGCTCATCTCCTCCAGC  4500

Qy  4501  AGTGCGTCCACGGCCGACCTGCCCATCGCCTTGACGGCTGTGTCAGGGGA  4560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4501  AGTGCGTCCACGGCCGACCTGCCCATCGCCTTGACGGCTGTGTCAGGGGA  4560

Qy  4561  GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCGACCACCGAGATGTCACCGGGA  4620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4561  GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTCGAAGCCGACCACCGAGATGTCACCGGGA  4620

Qy  4621  ACCGTGAGACCCCGCGGCGCCGGCCGCACGCGCCCGCCATCATGTCGCTGGCG  4680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4621  ACCGTGAGACCCCGCGGCGCCGGCCGCACGCGCCCGCCATCATGTCGCTGGCG  4680

Qy  4681  CACATGACGACGGCGGTGCAGCCCAGGTCGATCAGCGCGCGCGGCCGCCCCGGCCCCCTCC  4740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4681  CACATGACGACGGCGGTGCAGCCCAGGTCGATCAGCGCGCGCGGCCGCCCCGGCCCCCTCC  4740

Qy  4741  AGGGAGAACAGCGAGTGCTGCGCACGCGCCCCTGATCAGCGCGGCACACTCCCGGCACGTGC  4800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4741  AGGGAGAACAGCGAGTGCTGCGCACGCGCCCCTGATCAGCGCGGCACACTCCCGGCACGTGC  4800

Qy  4801  TCCCCACGCAGGCGGCCGCGCACGCGCTTCCGATCTTCCGCTGCACCGCGGAAGGGGGCGGCGC  4860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4801  TCCCCACGCAGGCGGCCGCGCACGCGCTTCCGATCTTCCGCTGCACCGCGGAAGGGGGCGGC  4860

Qy  4861  CCGACGGCGAGGGCGACGCGCTCGTGCCCCAGCTGCGTCGTGCCCACCGGGCCCAGGCGC  4920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4861  CCGACGGCGAGGGCGACGCGCTCGTGCCCCAGCTGCGTCGTGCCCACCGGGCCCAGGCGC  4920

Qy  4921  ATCGGCGGCCCGGCGTCGTCCGGGAGGCGCTCGATCCGGGGCGCTCGGGAGAACCCGGTTC  4980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4921  ATCGGCGGCCCGGCGTCGTCCGGGAGGACGAAGGGTGCCTCGATCCGGGGCGCGAGAACCCGGTTC  4980

Qy  4981  ACGAGGACGAAGGGCCACCTGCAGCCGGCCGCTGCGCCCATCACCTGGGGCGTCTCGGCGTG  5040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4981  ACGAGGACGAAGGGCCACCTGCAGCCGGCCGCTGCGCCCATCACCTGGGGCGTCTCGGCGTG  5040

Qy  5041  GTGTCCCGCGGGTGACCAGTCCGAGACGAAGATGCCGGACAACCGGTCCACGAGCATC  5100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5041  GTGTCCCGCGGGTGACCAGTCCGAGACGAAGATGCCGGACAACCGGTCCACGAGCATC  5100

Qy  5101  TCCGTGAGTTGTCGTCCGGGGTCTTCGCTCGGCGAGCGCCGCGCGGCCGAGCGGGCGTAG  5160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5101  TCCGTGAGTTGTCGTCCGGGGTCTTCGCTCGGCGAGCGCCGCGCGGCCGAGCGGGCGTAG  5160

Qy  5161  CCCTGACGCGTGAGCGCCTGCCCATCACCTGGGGCGGGAAGAGGGGTTGTCC  5220
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5161  CCCTGACGCGTGAGCGCCTGCCCATCACCTGGGGCGGGAAGAGGGGTTGTCC  5220

Qy  5221  AGTTCGGGGGTGACCAGTCCGACCTCGGCGCGGTCGTGCTCGGCGGTCGCTCGTAG  5280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5221  AGTTCGGGGGTGACCAGTCCGACCTCGGCGCGGTCGTGCTCGGCGGTCGCTCGTAG  5280

Qy  5281  CCCAGCGCGCTCCACCGCGTCAGCGCGTCCGGTGCCGCACCGGGTGCCGCGCGGCCGCCA  5340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5281  CCCAGCGCGCTCCACCGCGTCAGCGCGTCCGGTGCCGCACCGGGTGCCGCGCGGCCGCCA  5340
```

```
Qy  5341  CCGTTCAGCGACACCGGCTGACCGTGCCTTGCTGACCGCTGCCGTGCCGGCTGCGATGTCGGCG  5400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5341  CCGTTCAGCGACACCGGCTGACCGTGCCTTGCTGACCGCTGCCGTGCCGGCTGCGATGTCGGCG  5400

Qy  5401  AGCCGCATGGTCATGGCAACGCACTCTACCTGTCGGGGCGTCAGGGCGTGCCGCACCGCGC  5460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5401  AGCCGCATGGTCATGGCAACGCACTCTACCTGTCGGGGCGTGCCGCACCGCGC  5460

Qy  5461  GCGGAACCGGCGGACTGCGGGGCACGGCCGTCCCGCGCCACGACCACCGCGCCCGAAA  5520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5461  GCGGAACCGGCGGACTGCGGGGCACGGCCGTCCCGCGCCACGACCACCGCGCCCGAAA  5520

Qy  5521  CGATGGCTGAAAATGCTTGCAGCAAATTGCCGCCAACGTCTTTCGGCGGCTTTCGATCCT  5580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5521  CGATGGCTGAAAATGCTTGCAGCAAATTGCCGCCAACGTCTTTCGGCGGCTTTCGATCCT  5580

Qy  5581  GTTACGTTCCTGCAACCCGCGCGCCAGAGCCGTTGCGTGAGGCGTGCCAGACCT  5640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5581  GTTACGTTCCTGCAACCCGCGCGCCAGAAGCCGTTGCGTGAGGCGTGCCAGACCT  5640

Qy  5641  CCGCCCGATTCCGGGATCACCTCAGGGGAGTTCACAATGCGGCGTGGCATTGCGGGGCACCG  5700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5641  CCGCCCGATTCCGGGATCACCTCAGGGGAGTTCACAATGCGGCGTGGCATTGCGGGGCACCG  5700

Qy  5701  CGCTGTTCGCGGCTGTGGCCATGACGGCCATCATGTCGCGTTGGCG  5760
          |||||||||||||||||||||||||||||||||||||||||||||
Db  5701  CGCTGTTCGCGGCTGTGGCCATGACGGCCATCATGTCGCGTTGGCG  5760

Qy  5761  GCGGTACCGACGCGGCGGCCACCGTGTCCGGGCGCCCTCGCGCAGGGCTTCGAGGAAGGAGCACCCGA  5820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5761  GCGGTACCGACGCGGCGGCCACCGTGTCCGGGCGCCCTCGCGCAGGGCTTCGAGGAAGGAGCACCCGA  5820

Qy  5821  ACGAAGCCGAGAAGGCGACGTACCAGGCCATCAAGGCTTCGAGGAAGGAGCCGCCCGA  5880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5821  ACGAAGCCGAGAAGGCGACGTACCAGGCCATCAAGGCTTCGAGGAAGGAGCCGCCCGA  5880

Qy  5881  AGTCGACGTCAAGTCCGTGCCCCGACGTCAACGTCCGGCGGAGGCGTACGGCGGTCCGCG  5940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5881  AGTCGACGTCAAGTCCGTGCCCCGACGTGAATGCGTACGGAGGTGCGCCAAGTCAAGAACG  5940

Qy  5941  CCGCGGGGCGGCCAACTCGGCTCTTCTACAAGGAGAACTGCTGACGGAGGCCGGTGGTCGCGG  6000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5941  CCGCGGGGCGGCCAACTCGGCTCTTCTACAAGGAGAACTGCTGACGGAGGCCGGTGGTCGCGG  6000

Qy  6001  ACTTCGGCGACATCGGCTTCGGCGCCCAGGGTGCCGCACGCGCACGCGGCCGGGGGT  6060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6001  ACTTCGGCGACATCGGCTTCGGCGCCCAGGGTGCCGCACGCGCACGCGGCCGGGGGT  6060

Qy  6061  CGGACCACCTTCCCCAGGCGGCGAGCACCAGGGGAAGACCTACGCGGTCCCCGC  6120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6061  CGGACCACCTTCCCCAGGCGGCGAGCACCAGGGGAAGACCTACGCGGTCCCCGC  6120

Qy  6121  AGGTGATCGACACACCTGCGCTCTTCTACAACAAGGAACTGCTGACGAAGGCCGTCG  6180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6121  AGGTGATCGACACACCTGGTCTCTTCTACAACAAGGAACTGCTGACGAAGGCCGTGTCG  6180

Qy  6181  AGGTGCCGGGGCTCCCGCCTCCCGGAGTCGAGCTGACGAAGGCCGCCGAGAAGACCG  6240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6181  AGGTGCCGGGGCTCCCGCCTCCCGGAGTCGAGCTGACGAAGGCCGCCGAGAAGACCG  6240

Qy  6301  GCGGAGCGGCGCCTCTACTCGGGCCCAGTGCGGGAAGAACACCGTGCCCTACCTCTACG  6360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6301  GGGAGGGGCGGCGCCTCTACTCGGGCCCAGTGCGGGAAGAACACCGTGCCCTACCTCTACG  6360

Qy  6301  GGGAGGGGCGGCGGACCTCGGCGGCGACCTGGTTCGACGAGAAGAACACCGTGTCG  6360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6301  GGGAGGGGCGGCGGACCTCGGCGGCGACCTGGTTCGACGAGAAGAACACCGTGTCG  6360

Qy  6361  GTGTTGCGGCCTACCGCGTCATCAAGGACCTCGTGGACACGCAAGGCCGCCATCACCGGACG  6420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6361  GTGTTGCGGCCTACCGCGTCATCAAGGACCTCGTGGACACGCAAGGCCGCCATCACCGGACG  6420
```

```
Qy  6421 CGTCCGACGCGCTGAACAACATGAGAACGCCTTCAAGTCGGGCAAGTCGCCATGATGG 6480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6421 CGTCCGACGCGCTGAACAACATGAGAACGCCTTCAAGTCGGGCAAGTCGCCATGATGG 6480

Qy  6481 TCAACGGCCCCTGGGCGTCAAGGCGTCAAGGCGGGAGCCGCTTCAAGGACGGCCGGCA 6540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6481 TCAACGGCCCCTGGGCCATCGAAGGGGAGCCGCTTCAAGGACGGCCGGCA 6540

Qy  6541 ACTTGGGGGTCGCCCCCGGGTCGCCGGACAGGGCTCTCCCCAGGGCGGGT 6600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6541 ACTTGGGGGTCGCCCCCGGGTCGCCGGACAGGGCTCTCCCCAGGGCGGGT 6600

Qy  6601 GGAACCTCTCGGTACGGCGGGTCTGAAGAACCTCGACGCCTTCGTTGAAGT 6660
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6601 GGAACCTCTCGGTACGGCGGGTCTGAAGAACCTCGACGCCTTCGTTGAAGT 6660

Qy  6661 ACATGAGCTCCGCCAAGGTGCAGCAGCAGACCACCGAGAGCTGAGCCTGCCCACCC 6720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6661 ACATGAGCTCCGCCAAGGTGCAGCAGCAGACCACCGAGAGCTGAGCCTGCCCACCC 6720

Qy  6721 GCACGTCCGTCTACCAGTCCGTCCGTCGCGGACAAACGAGATGGTGAAGTTCTTCAAGC 6780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6721 GCACGTCCGTCTACCAGTCCGTCCGTCGCGGACAAACGAGATGGTGAAGTTCTTCAAGC 6780

Qy  6781 CGGCCGTCGACAAGGCCGTCGAACGGCCGTCGGATCGCCGAGGGCAATGCCCTCTTCGAGC 6840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6781 CGGCCGTCGACAAGGCCGTCGAACGGCCGTCGGATCGCCGAGGGCAATGCCCTCTTCGAGC 6840

Qy  6841 CGATCCGGCTGCAG 6854
         ||||||||||||||
Db  6841 CGATCCGGCTGCAG 6854
```

RESULT 2
US-09-988-384B-3/c
; Sequence 3, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE: 2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 3
;  LENGTH: 12441
;  TYPE: DNA
;  ORGANISM: Streptomyces venezuelae
US-09-988-384B-3

  Query Match             7.7%;  Score 527.4;  DB 9;  Length 12441;
  Best Local Similarity   53.3%;  Pred. No. 1.2e-98;
  Matches 1475;    Conservative    0;  Mismatches 1181;    Indels  109;  Gaps  13;

```
Qy  2449 GCGA--ACTCACGGCGACGGCACCGCGTTCGGGTTACGCATCACCTACCTCCACCAGC 2505
         ||     ||   ||    ||   |||||  |||| || |  |||| || ||||||||||
Db  9121 TCCAGTCGCTTCTGCGGAAACGGCAGCACTCGGCCGCTTCCTGGCGACACG 9062

Qy  2506 CCCGCCGCTCGGTCTCGTCGCGCAGGACGAGGCCGCTCAAGGCGTCAAGGACACGCCGCCA 2565
         ||   |   ||| ||| ||  ||||||   || |||||||||||||||||||||||||||||
Db  9061 AAGAGCCCGCAGAATCGCGGACGACGCACTTCTCGTCGGAGACCGAGACGCGCGACGACA 9002

Qy  2566 ACTTCCTGTGTACCTGGGGACAACTACCTGCCC---CAGGCGTCACCGACTTCGCGC 2622
         ||     |||| ||| ||||| ||| ||  || |   ||| | ||| |  |  ||||| 
Db  9001 CCTGCCCCTGATCCTGGGCGACAAACATCTTCCACACGGGCGCCCTCTACACGCTCCTGC 8942

Qy  2623 GCCAATCGGCCGCCGATCCCGGCGGCGGCTGTCGTCACCCGGTGCGCGGACCCGT 2682
         ||  |  ||  ||| |  |  |||||| |  |  ||||||||||||||||||||||
Db  8941 GGGACAGCATCGCGCCGCTCGACGGCGTCGCGTGTCTTCGCGTGTCTCCGTCAAGGACCCCG 8882

Qy  2683 CCGCCTTCGGGTCGGGAGGTCGACGGCGGAACGTCGGCGGGCACCTCGTCGAGAAAC 2742
         ||    | ||||||  || ||| ||||||   | ||  ||||| || |||||||||||
Db  8881 AGCGGTACGGCGTCGCCGAGGTTGGACGGCGAGGCGCGCGGCTGGACCTCGTCGAGAAGC 8822

Qy  2743 CCGACGTCCGCGCAGCTCGCTCGCTGCGTGTACGCGTGTGTACGCCCGCCGCGTCC 2802
         ||| |  || | |||   |  |  ||  |||  ||  || |||||| ||  ||||||
Db  8821 CCGTCAAGCGGCCGCTCTCAACCTCCACCGTCCGCCCGCCGGCGCGGCGGAGGCTGCA 8762

Qy  2803 ACGAGGCGTACGGCGTACCACCCCCTGGATCGCGGAGCTGGAGATCACCCACCCCCG 2862
         ||||||| ||||||  || | |    |||||  | | ||||||||| || |||||||
Db  8761 TCGACATCGCCAAGAACATCCGGCCCTCGCCCTCGGCGCGCGGAGCTGGAGATCGTCA 8702

Qy  2863 TGCAGTCGATGATCGACCACCGGGGCCGCTGCCCGGGGCCGA--GACCACCACCCGGCCCT 2919
         || ||| ||| | | || ||||  | ||  ||  |||  |  |  |||  || || ||||||
Db  8701 ACCGCGTCTACCTGGACGCCACCGGGCGGGCCGGCGAACTCGTCAACCTGGGCCCGCTTCGCCT 8642

Qy  2920 GGCGCGGACACCGCAGCGGCGGAGGACACTGCTCGAGGTCAACGTCAGCTCCTGACGGAC 2979
         ||||||||||||||||||||  |||||   |||| |||||||||||||||||||||||||
Db  8641 GGCTGGACCACCGCAGCGGCCCACGACTCTCGGTCCCAGGTCCGGTCCGG 8582

Qy  2980 TGGAGGGCCGCATCGAGGGGAAGGTCGACGCCGAGGTCGGTCGGCACGCTGGTCGGGTCGGGTCCGG 3039
         ||                                ------GCAGGGCTCTGCGATCGCGG
Db  8581 AGGAGCG                           GCAGGGCTCTGCGATCGCGG 8555

Qy  3040 TGGCCCAAGGGCGCGATCGTGGCGGGGTCACACGTGGTGGCCCGGTGCCGGTGATCGGCGCGG 3099
         ||     |||  |   | |   ||||  |   | | || |||||  || |||| ||  |||
Db  8554 GCCTTGAGGGAGATCGCCTTCCGCATGGGCTTCATCGACGGCGGCCTGTCACGGCCTGG 8495

Qy  3100 GTGCCGTCGTCAGCAACTCCAGTCGGCCGTCGTCACCCGTCACACCTCCATCGGGCGGAGGACTGCCGGG 3159
         ||     || |  ||    |  | | |||  | ||| |  |  || |||||||| | ||
Db  8494 GAGAAGGCCTCTTCCCGACCGAGTACGGCAGCTA------TCTGATGGAGATCGCCGG 8442

Qy  3160 TCGAGGACAGCGGCCATCGAGTACTCCGTCCTGCTGCCGGCCCACTCGCCGGGGGGCGT 3219
         |       | | ||  ||  ||   ||| | ||||| |||  ||    |||||||||||
Db  8441 CGCGAGGGGAGCCCCGTTCCGAGGGCACCTCGGCCGTTCCCACGACCGAGCAGCGGCCA 8382

Qy  3220 CCCGCATCGAGGCGTCCTCATCGGCGGACCGGGCGTTCCCACCCGCGCCCCCCGTCTCC 3279
         |     || |||  | |  |  | ||||  ||| |||| || |   | |||||| -------
Db  8381 CCGACAGTGCGACCCCGACCCCACCGCGCACCGCACCGACAGTGCGACC---------- 8332

Qy  3280 CGCAGGGCTCACCGACTGTGATCGGCGACCAAGGTGTATCTCACCCATGACCAC 3339
         |  | |||                          |||      ------ACGGCAGT
Db  8331 ------CACACCCGCGACCTACCGACCGAGAAGGAG-----------ACGGCAGT 8291

Qy  3340 GACCCATCCTCGTCACCGGCGAGCCGGTTCACCGGCTTCATTCGCTCCGGCCTCCT 3399
         ||     |||| |    ||| |||||| |    ||||||||  ||| ||| ||||||
Db  8290 GCGGGGTTCTGGTACCGGAGGTGCGGCGTGCTCATCGGCTCATCCGGCGGCAGCTCCT 8231

Qy  3400 GTCGCCAAGCAGCTGCTCCCCCCCGGGGCCCCG------GCGGCGTCCGGCGCCCTCGACAAACTCACCTA 3453
         || |  |  ||| |  | | ||||  ----- |||||  | || || || |  ||||||||||
Db  8230 CGCCCGGGGGGGTACCCCGGCGGTGCCCGCCGATGAGGTGGCGGCCTCCTGGACAGCCTCACCTA 8171

Qy  3454 CGCCGCAGCCGTCGTCCGGGGTGTCCGGGAGCGGCGGCGTCGTCGTGCTGCACCCGGCTTCGTTCCA 3513
         |  |  |  | |   |  |    | ||| ||  ||| || || ||| ||| |||  |||  || |||
Db  8170 CCGGCGGCAGCCGCCAAACCTCGCCGGGTGTCTGGCGGCGTGACTCGGACGCATCGT 8111

Qy  3514 GGGCGACGTGTGCGACACCGGCTCGTGCGACGACGACATCGT 3573
         ||| || ||  || |  |||| || ||| |  |  ||||||||
Db  9181 TCATGCTCGGCGGTATTCGCGGAGATTCAAATCATCTGACCCCCCAGGACGACATCGT 9122
```

RESULT 3
US-09-860-846-3/c
; Sequence 3, Application US/09860846
; Patent No. US2002016474A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-3

Query Match            7.7%;  Score 527.4;  DB 9;  Length 13613;
Best Local Similarity  53.3%;  Pred. No. 1.2e-98;
Matches 1475;  Conservative 0;  Mismatches 1181;  Indels 109;  Gaps 13;

Nucleotide sequence alignment between Qy (query) and Db (database) sequences, shown in paired rows with position numbers and match bars (|).

```
Db  9748 AAGAGCCCGCAGGAATCGCGACGCACTTCTCGTCGGAGCCGAGCACATCGGCGACGACA 9689
Qy  2566 ACTTCCTGCTGTACCTGGGGACAACTACCTGCCC---CAGGGCGTCACCGACTTCGCC 2622
              || |||  ||  ||||| ||    ||||| |-||||| |- |||| ||||| ||||
Db  9688 CCTGCGCCCTGATCGTCTGGGCACAACATCTTCCACGGGCGTCGCCGCCTCCTACACGCTCCTGC 9629
Qy  2623 GCCAATCGGCCCGCGATCCCGGCGGCCCGGCTGCTGCTCTCCCCGTGCGGACACCCGT 2682
              |||| ||||||||  ||| | ||||| ||| ||| ||| ||||||  ||||| ||
Db  9628 GGGACAGCATCGCGGCCCTCGGCGACGGGAACGTGCTGCGCCTTGGAGGAGAAAC 9569
Qy  2683 CCGGCCTTCGGCGTCGGCGACGGGACAACGTGCTGCGCCTTGGAGGAGAAAC 2742
Db  9568 AGCGGTACGGCGTCGCCGAGGTGGACGGCGACGGGCCGGGCTGACCGACCTGTCGAGAAGC 9509
Qy  2743 CCGACGCCCGGCAGCTCGCTCGCCGCCCATCGGCGTACGGCCTCTCAGCCCGGCGTCC 2802
              |||| ||   ||  ||||  |||| ||  ||||  ||  ||| ||-| |||| ||
Db  9508 CCGTCAAGCCGCGCTCCAACCTCGCCGCCGCCGCCGCCGCCGCCATCGGCCGGCCCG 9449
Qy  2803 ACGAGGCGGTACGGCCATCACCCCGTCCGCGCGGCGAGCTGGAGATCACCCACGCCG 2862
Db  9448 TCGACATCGCCAAGAACATCCGGCCCTCGCCGCCGCCGAGCATCACCGACGCTCA 9389
Qy  2863 TGCAGTGGATGATGACCGGCCGATGCCTGCGCGTACGGCCGA---GACCACCACCGCCCCT 2919
               || |||||||||||||  || ||| ||||   |||||||   |||||||  || ||||
Db  9388 ACCGGCGTCACCTGAGCGGCGGCCGGCCGAACTCGTCACCGTCGGTCAACCTGGCGCCGCCCGGAC 9329
Qy  2920 GGCGCGACACCGCACCGGCACCCCACGACTCGCTCCTGCGCGACGCCAGTCGGGTCCTGG 2979
               |||||||  ||||  ||| |||||| |||||   ||||    |||| |||| |||
Db  9328 GGCTGGACACCGCACCCCACCGACTCGCTCCTGCGCGGGTTGGCCGGTCCGGGG 9269
Qy  2980 TGGAGGGCCGCATCGATCGGAGGGGAGGTCGACGCGCCACAGCCTGGTCGTCGGGTCGGGG 3039
               | ||  || ||| || |||      | ||||  ||||||  || |||||||||||
Db  9268 AGGACG------------------------GCAGGGCGGCTGTTTGGATCGGCGG 9242
Qy  3040 TGGCCGAAGGCGGCGATCGTGCGGGGGTCACACGTGGTGGCCCGGGCCGGCCTGG 3099
               || |||    || |  |   |      || ||| |  ||||   ||||||  || |
Db  9241 GCCTTGAGGAGATCGCCTTCCCGCATGGGCTTCGGGCCCGTACACCTCCATCGGGGAGGACTGCCGGG 9182
Qy  3100 GTGCGCGTCTCAGCAACTCCAGTGTGCGGCCTGGCGCGGTACACCTCCATCGGGGAGGACTGCCGGG 3159
               ||||||  ||   ||||| ||||  || || ||||||||||||||||||||||
Db  9181 GAGAAGGGCCTCTCCGCACCGACTACGGCCAGCTA------TCTGATGGAGATCGCCCGGC 9129
Qy  3160 TCGAGGACAGCGGCCATCGAGTACTCCGGTCCTGCTGCCGGCGGCGGCGGTCGGAGGGGCGT 3219
               |   |||  ||| |||||||   | |     |||||||| |||||||||| |||| |
Db  9128 CGCGAGGAGGCCCGTGAGGGCACCTCGCGGCCGCACCGCGTTCCCACGACCGACGCGCCCA 9069
Qy  3220 CCCGCATCGAGGGCGTCCCCATCGGCGCGCGCGTCGTCGGCGGCCGCGCCCGTCCC------ 3279
Db  9068 CCGACAGTGCGACCCGGCGCACCGCGACCGCACCGACGAGTGCGACC-----------ACGGCAGT 9019
Qy  3280 CGGAGGCTCACCGACTGGTGATCGGCGACCAGCAAGGTGTATCTCACCCCATGACCAC 3339
               ||||||  ||   |||||      |||||||      |-||||||||   ||||||||
Db  9018 ------CACCCGGCGACCTACAGCGCGACCGAAAGGAAG------ACGGCAGT 8978
Qy  3340 GACCATCCTCGTCACCGGCGAGGTGCGGGCTTCATTCGGCGGCTTCATTCGTCGCGGCTCCT 3399
               |||| || | ||||   ||||||||||| ||||||||  ||||||||||||  ||||||||||
Db  8977 GCGGCGTTCGTGGTGACCGAGGTGCGGGCCGTCGCACTTCGTCGCGGCAGCTCCT 8918
Qy  3400 GTCGCCGGGGCCCCG------GCGGCGTCGCGTGACCGTGGAGTCCC 3453
               | ||||||||||||      | |||| ||  |||| ||||| |||||
Db  8917 CGCCGGGGGCTACCCGCGACGTACCCGGGCCGGTGATCGTCTCCTGGACAGCCGTCGTCGTCACCTA 8858
Qy  3454 CGCGGCCAGCGTCGGTGTCCGGGAGGCGGTCGCTGTCGTGACCATCCGGCCTCCACCTTCGTCCA 3513
               ||||||  || |||| ||||| ||||| |||| || | |||||||||   |||||| |||||||
Db  8857 CGCGGGCAACCGCGCGCAACCTGACCTGGGGACCCGGGTGGACCCGCGACCTGCGCGTTCGTCCA 8798
Qy  3514 GGGGGACGTGTGCGACGCTGGTCGTCTGCGCCGTCTGCGACACCGCACGCGACATCGT 3573
               ||||| | |  | | | |  |||||  ||  || || ||| ||| |||||| ||||||||
Db  8797 CGGGGACAATCCGGCGACGCGGCGGCCTCCTGCGCGTCGACGTATGCGGCCATCGT 8738
Qy  3574 GCACTTCGGCCGCGAGTCGCCGAGGACACGTGACGCCGGCTTCCCTGGGCGACGACG 3633
               |||||||||||||   | |  |||||||||| || ||||||||||||||||||||
Db  8737 CCACTTCGGCCGGGAGAGCCACGTGGACCGTGGACGCGGCGCCGTCGTTCACCGGA 8678
```

```
Db  7657 GCGACCCCGTCGCCGTCGCCGTCGAGCGCGACGAGGACCACCCACCCTGGACCGCTGCTCGTC 7598
             ||||     ||| |||||||     | |||||||  |||| ||  |||||||| |||
Qy  4693 GTTGCAGCCGCCAGGTCGATCAGCCCCCGCACCCCGCGCGG-CGGCCTGGCCCCCCTCAGGGAGAACAG 4751
Db  7597 GAGAAGGCGATCACCCCCGCACCCGGCGCGTCCTCCCCGCCGCACTCCAGGCTCACGGGCCACCCC 7538
Qy  4752 CGAGTGCTGCACGAGCTCCCGGGACTCCCGCCGCCGCAGTGCTCCCGCACGCGCC 4811
Db  7537 GCCGACCATGGACGCCCCCTCGAGCGACGCCGACGAGCCGGCCGCAGCAAGGGCGCGGCGAC 7478
Qy  4812 GGCCGGGAACCCCTCGAGCATCCGGTCGCCGCCGACGGGCCCGGATCGGCGCGAG 4871
             |||||||||| |||| | || | |||| ||   || ||| ||||||||| |||||
Db  7477 GCCGCGCAGGCCCACG---GGCGCCCGTACCGGCGGCGGCGATCGGCGCAGGCGCGGGGTCG 7421
Qy  4872 GCCGACGCGCTCGTGCCCCAGTCCGCCAGTGCTGCGCGCACGGCGCCAGGGCGCGCATCGCGGCCCCG 4931
Db  7420 GTGGCCGCGTTCAGCTTCTACCCGGGCAAGAACCTCGGGCTGCTTCGGCGTCCGGCGCGGACGAA 7361
Qy  4932 GTCGTCGGGGGAGGACGAAGGGTGCCTCGATCCGGGGCGAGAACCCGTTCACGGAGGACGAA 4991
             |||||||| ||||| |      || | | |||||||| ||| |||| ||||||||||||
Db  7360 GTCGTCACCGGCGCCGGCTCGGCGCTCGCCGGATCGTCCGGATCGCTCCCCAACTACGGCTCG 7301
Qy  4992 GGGCA 4996
             |||||
Db  7300 CGGCA 7296
```

```
Query Match          7.7%;  Score 527.4;  DB 9;  Length 13613;
Best Local Similarity 53.3%;  Pred. No. 1.2e-98;
Matches 1475;  Conservative 0;  Mismatches 1181;  Indels 109;  Gaps 13;

Qy  2269 TGAAGGCCCCTGGTTCTGGCAGGTGGAACGGCAGCAGACTGAGGCCGGTCACCCACACCG 2328
             |||||||||||||| || |||||  |  |  ||||||  ||||||  ||||||||||||
Db  9988 TGAAGGGAATAGTCCTGGCAGGTGGAACGGCAGCAGACTCGGCCGGTCACGCGGGTCA 9929
Qy  2329 CCGCCAAGCAGCTGCTCCCCATCGGCCGTGCCTCTTCTACGGCGTCGGAGTCCC 2388
             || ||||||| | |   || | ||||| ||  || |||||||| |||||||| |
Db  9928 TTTCGAAGCAGATTCTTCGGTGTCGGGCGAGGCGGTCGGCGTACGGCGGGGAGATCC 9869
Qy  2389 TCGCCGCGGCGGGGTCCGGTATTCGGCGACGGGCCGCGGGTATTCGGCGGCGAGTCGCCGGGGGATCC 2448
             |  |||||||||  || || |||||||   |||||    || ||||||||||||||||||||
Db  8857 CGCGGGCAACCGCGCGCAACCTGACCTGGGGACCCGGGTGGACCCGCGACTGCGCGTTCGTCCA 8798
Qy  2449 GCGA---ACTCACCGGCGACGGCACCGCGGTTACGGCCATCACCTTCCACCAGC 2505
             ||||    ||||| |||||||||   |  ||| |  ||||| |||||| ||||
Db  8797 CGGGGACAATCCGGCGACGCGGGCGGCCTCCTGCGCGTCGACGTATGCGGCCATCGT 8738
Qy  9808 TCCAGTTCGCTTCTTCCGGAAACGGCACCTGGAATAGAACTCGACTATGCGGTCCAGA 9749
             ||| | |  |||  |  ||||   |||||  |  || ||||||  || ||||| |||
Db  2506 CCGCGCCCCGTCGCGGTCTCGCGGCTGCGACCGGTGCGCCACGAGCCACGTGACGCG 2565
             ||  |||||||||||||||
```

```
Qy   4693  GTGCAGCCCAGTGCGATCAGCGCGACGCGG-CGGCTGCTGCCCCCTCCAGGGAGAACAG  4751
           |||  ||| ||||  |||||   |  ||  ||   |||  | | |||   |||
Db   7597  GAGAAGGCGATCACCCCCGCACCCGGGCTCCTCCGCCGCACTCCAGGTGCTCCCGCACCCC  7538

Qy   4752  CGAGTGCTGCACGAGCTTCCTCGGACTCCCGCCGCACTCGCCGTGCTCTCCGGGCTCCGC  4811
           ||| || ||||||  |||||||||||| |||  |  || |  | ||| |||| |||||
Db   7537  GCCGACATGGACGCCCTCCTCCGCGGAGCTCGCGGAGCTCGCACATCGTCGGAGGAC  7478

Qy   4812  GGCCCGGAACCCCTGATCTTCCGCTCGAAGCGGCCTCGGACCTCAT  4871
           ||||||||  || |  | ||||| ||  |||||
Db   7477  GCCGCCAGGAGGCCCACG---GCGCCCGGTACCCGGGCGATCGGCGCCGGGTCGTCG  7421

Qy   4872  GCCGAGCGCGTCGTGCCCCAGCTCCGCCAGTGCGCCAGGCGGCATCGCGGCGCCG  4931
           | |||  ||||  |  | |||| ||| ||||
Db   7420  GTGGCCCGGTTCAGCTTCACCCGGCAAGAACCTCGGCTGCTGTTCGGCGGACGAAC  7361

Qy   4932  GTCGTCCGGGAGACGAAGGGTGCCTCGATCCGGGGCGGAGAACCCGTTCACGGACGCGAA  4991
           ||||| |||||   ||| | |  |  |  ||| ||| ||| ||||
Db   7360  GTCGTCACCGGCGACCCCGAGCTCGCCGAACGGCCTCCGGATGCTCCGCAACTACGGCTCG  7301

Qy   4992  GGGCA  4996
           |||||
Db   7300  CGGCA  7296
```

RESULT 5
US-09-861-289-3/c
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;   APPLICANT: Sherman, D.H.
;   APPLICANT: Liu, H.
;   APPLICANT: Xue, Y.
;   APPLICANT: Zhao, L.
;   TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;   FILE REFERENCE: 600.438US1
;   CURRENT APPLICATION NUMBER: US/09/861,289
;   CURRENT FILING DATE: 2001-05-18
;   PRIOR APPLICATION NUMBER: 09/105,537
;   PRIOR FILING DATE: 1998-06-26
;   NUMBER OF SEQ ID NOS: 43
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 13613
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-3

Query Match             7.7%;  Score 527.4;  DB 10;  Length 13613;
Best Local Similarity  53.3%;  Pred. No. 1.2e-98;
Matches 1475; Conservative    0;  Mismatches 1181;  Indels 109;  Gaps  13;

```
Qy   2269  TGAAGGCCCTGCCTGGCAGTGGAACCGGCCAGCAGACTGAGGCCGTTCACCCACACCG  2328
           |||||  | ||| |||||  |||||||  ||  || |||||  |  ||||  ||||
Db   9988  TGAAGGGAATAGTCCTGGCCGGAGCGGAACTCGGCTGCATCCGGCGACCTCGGTCA  9929

Qy   2329  CCGGCCAAGCAGCTGCTCCCCATGCCAACAAGCCCGTGTCTCTTCTACGGCGCTGGAGTCCC  2388
           |  |||||| || || |||| ||  | | |||||   || ||  || ||||
Db   9928  TTTCGGAAGCAGATTCTTCGGTCTACAACAAACCGATGATCTACTACTACCGCTGTCGCGTTC  9869

Qy   2389  TCGCCGCGGCGGGTGTCCGGGAGGCCGGCGTCGTCGGGCCGTACGGCGGGAGATCC  2448
           |  | |  |||||    ||  |    | ||||   ||||| |   || || |
Db   9868  TCATGCTCGGCGGTATTCGGCGGAGATTCAAATCATCTCGACCCCCAGCACATGGAACTCT  9809

Qy   2449  GCGA---ACTCACCGCGACGGCCGTTCGGGTTACGCGTCATCACCTACCTTCCACCAGC  2505
           ||||    | ||||| ||   ||  |  | ||||  ||  |  | | ||||||||
Db   9808  TCCAGTTCGCTTCTCGAAAACGGCAGGCACCTGGGAATAGAACTCGACTATGCGGTCCAGA  9749

Qy   2506  CCCGGCCCCGCTCCGTCTCGCCACGCCGGTGCGCATCGGCGCTTCCTGGGCGACGACGCG  2565
           | | ||||  ||  || ||||  || | ||  | | || ||   ||||||||||||||
Db   9748  AAGAGCCCGCAGGAATCGCCGGACGCCACTTTCGTCGTCGGCGACCATCGGCGACGACA  9689

Qy   2566  ACTTCCTGCTGTACCTGGGGGACAACTACCTGCCC---CAGGGCGTCACCGACTTCGCCC  2622
```

```
Qy   3634  CACCAACGTGCTGGGCACCCAGTTCTCGCTCGACGCGCCGTCGTCGCCAC  3693
           |||||||||||||| ||||||  |||   | |  ||    |||  |
Db   8677  GACCAACGTGCAGGGCACGCAGACGTGCTGCCAGTGCGCTGCACGCGCACGGGGCGTCGGCCG  8618

Qy   3694  CTTCGTGCACGTCTTCCACCGACGAGGTGTACGGCGCCGCCGACCGCGCGGGA  3753
           |||||||||||||||||| |||||||| | | ||| || ||  ||  |
Db   8617  GGTCGTGCACGTCCACCACCGAGGTGGTCGATGCACTCCGGCTCTGGACCGA  8558

Qy   3754  GAGGACCCCTGCTTCCGACCTCGCCGTCGAAGCGGCCTCGGGACCTCAT  3813
           ||||  | |||  ||  | ||  |  |  || |||||
Db   8557  GAGCAGCCCGCTGGAGCCCAACTCGCCCTACGCGCGGTCCAAGGCGGCTCCGACCTCGT  8498

Qy   3814  GGGGCTCGCCCACCACCGACCACCAGGCTCCGGTGACGCCTCCGGTGACCGGTGTTCGAACAA  3873
           ||||| |||| |  |  |||  ||     |||| |   | ||
Db   8497  TGCCCGGCTACCACCGACGTACGGCCTCGACGATCACCCGCTGCTGCAACAA  8438

Qy   3874  CTTCGGCCCCACCAGCATCCCGAGAAGCTCATACCGGCCTTCCTGACCAGCCTCCTGTC  3933
           || ||| ||||| ||  ||||||||||||  ||  | |  || | || ||
Db   8437  CTACGGCCGTACCAGCACCCCGAGAAGCTCATCCCCCTTCGTGACGAACCTCCTGA  8378

Qy   3934  CGGCGGCCACCGTTCCCCTCTACGGCGCACGGTCGCGCACGTGCTGCCACGTCGA  3993
           |||||||  ||||||  ||     | | ||||  | |  |  ||| || | |
Db   8377  CGGCGGGACGCGTCCCGTGTACGGCGACGGCGCGGAACGTCGCGGAGTGGTGCACACCGA  8318

Qy   3994  CGACCACGTTCAGGGCCGTCGAACTCGTCGGCGTGTCGGCGGTGCGGAGAGAATCTACAA  4053
           ||||||  |  |  ||   |||  |  |||| |  ||| |||   ||| ||||||| |||
Db   8317  CGACCACTGCCCGGGCCATCGCGTCGTCGTCGGCCGCCACCGTCTCGCGGCGGAGATCTACCA  8258

Qy   4054  CATCGGGGGCGGCACCTCGCTCGCCAACCTGGAGTCACCCGTTGCTCGCACTGTG  4113
           ||||| ||||  || || |||   | |||    ||| || | |||| | | |
Db   8257  CATCGGCGGCGGCCTGGAGCTGACCAACCGCCGACTGCGCCGAACTCACCGGACTCGCT  8198

Qy   4114  CGGCGGCGGCCCGGAGCGCATCGTCCAGTCAGTGAAACCGGCAAGGGCACGACCGGCGCTA  4173
           |||| ||  |||  ||    ||| | || |  | | || |  | || | ||   || || |
Db   8197  CGGGCCGACTGTCCTCGGTCCGGAAGGTCGCCCACCGGGGCCACCACCTCGCGCTA  8138

Qy   4174  CGCGGTCGACCAAGCAGATCACCGCGGAACTCGGTTACCGGCCCGCCGACCTTCGC  4233
           | | | |||| | | | |  | |||  ||||| ||| |||
Db   8137  CTCCTCGACGGCCGGCGGCGAGATCGAGCGCGAGGCGGCTCGAGCTACCCCGAGGCCGACTCCTTCGC  8078

Qy   4234  GACCCGCTGGCCGACACCGGAAGTGTACGAGCGGCACGGAACCGCGTCGGTGGCCGCT  4293
           | || |  |||  ||||    ||| | |||||||||||  | ||| ||||||||
Db   8077  GGACGGCCTCGGACGGCAACCGCTCCGGTACCGGGAGAACCGCGTCGGTGGGGAGCCGCT  8018

Qy   4294  GCTCGCCCGGACATGACGTCGG------GCCGACCGCAACCACCGGCCCCGCCGG  4344
           | | || |  ||  | |||       |||| |  |    ||||  |  |||
Db   8017  CAAGGCGACACCGCCCCGCAGCTGCCCGCCACGCGTGTCCGTGAGGCGCAGCCGC  7958

Qy   4345  CACACCGGCGCGCCCGCGCCCGCCGGCCGGCGGCGCCGGTCGGTGCGCCGGTCAGCGCCGGCCGG  4404
           |  |||||  |||   ||||||| ||    |||  |  ||  |||| || |  ||
Db   7957  GCCGAGACCCCCCGCTGCGTCCCCGCCCTTCCTCGACCTCAAGGCCCTACGAGGAGCTCCGGCGCG  7898

Qy   4405  CCCCGGCGGCGGCGCGGTGACCGGGTGTGACCCGGTCCATGAAGACGAA  4460
           |||||  ||||||||||| |  ||| |   ||| ||  |  |   ||||
Db   7897  GAGACGAACGCCGACGCCGGATCGCCCGGGTCCTCCTCGGGGCGCTACCTCCGGACCCGAA  7838

Qy   4461  TTCGGCGGGGCGGGGCGGCGTTCGGCTCATCTCCTCCAGCAGTGCGTCCACGGCGACCTG  4520
           || | |  ||   ||| || |  |    |  |    || ||   || ||
Db   7837  CTCGAAGGATTCGAGGCGGAGTTCGCCGCGCTACTGCGCGAGACGACCACGCCGTCGGCGTG  7778

Qy   4521  CCCATCGCCTTGACGGGCT------GTCTGATGGTGGTCAGGGAGGGTCGGTGAAG  4572
           |||  ||||    ||||  |  |  |  |  |  |  || |
Db   7777  AACAGCGGGATGGACGACCCTCCAGCTGGCATCGGACCCCGGGGAC  7718

Qy   4573  GCCATGAGCGGCGAGTCGTCGAAAGCGACCACCAGGATGCGTCCGGAACCGTGAGACCC  4632
           ||  ||  || | || | |  ||    | | |||  ||| ||  ||  ||| | || |
Db   7717  GAGGTGATCGTCCCCTGACACACGTACATGCCCTGCCGACTACAGCGTTCGCCGCCACCGGC  7658

Qy   4633  CGCCGGCCCCGCTCCGTCTCGCCACGCCGGGCCATCATGTCGCTGCCGCACGCGCGCATGACGCGG  4692
           ||| || || |  |  || | |  | ||  |   ||||||||||||
Db   7657  GCGACCCCGTGCCGCTCCCGCTGCCCGCTCGAGCCCGCAGAGGACGCCACCCACCCTGCCCCTGCTGCCGTC  7598
```

us-09-922-683-7.rnpb

Thu May 29 15:11:30 2003

```
Qy  3352 CACCGGCGGAGGGGCTTCATTCGGTCGCCTCCGCCTACGTCCGTCTCCGTCTCCGCGGGGC 3411
         | ||||||||||||  ||||| ||||| ||||||  | |||  || ||||| ||  |||||||
Db     1 CCCGGGCGGGGCGGGGGGTTCATCGGTCGCCTCGCGGCTCCGGCAGCCTCGCCGTCCCGGGGGC 60

Qy  3412 CCCGGCGGGCGTCGCGGTGACCGTCTCTGACAAACTCACCAGCGGCGCCGAGCTCGCCCGAG 3471
         ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db    61 CCCGGGCGGGCGTCGCGGTGACCGTCCTCGACAAACTCACCACCGGCGCCGAGCTCGCCCCG 120

Qy  3472 CCTGCACGGCGGTGCGTGACCATCCGGCCTCACCTTGTCCAGGGGACGTGTGCGACAC 3531
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCTGCACGGCGGTGCGTGACCATCCGGCCTCACCTTCGTCCAGGGGACGTGTGCGACAC 180

Qy  3532 CGCGCTCGTCGACACGCTGGCCGCGCACGACGACACATCGTCGCACTTCGCGGCCGGAGTC 3591
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CGCGCTCGTCGACACGCTGGCCGCGCACGACGACACATCGTCGCACTTCGCGGCCGGAGTC 240

Qy  3592 GCACGTGCACCGCTCCATCACCGACAGCGGTGCCTTCACCCGACCACGTGCTGGGCAC 3651
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCACGTCGACCGCTCCATCACCGACAGCGGTGCCTTCACCCGACCACGTGCTGGGCAC 300

Qy  3652 CCAGGTCCTGCTCGACGCGCCGCTCCGACGCGCCGCTCCGCACGGTGCGCACCTTCGTGCACGTCTCCAC 3711
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCAGGTCCTGCTCGACGCGCCGCTCCGCACGGTGCGCACCTTCGTGCACGTCTCCAC 360

Qy  3712 CGACGAGGTGTACGGCCTCCCTCCCGCACGGGCGGCCCGCGGAGAGGACCCCCTGCTTCC 3771
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CGACGAGGTGTACGGCTCCCTCCCGCACGGGCGGAGCGGCCCGCGGAGAGGACCCCCTGCTCTCCC 420

Qy  3772 GACCTCGCCGTACGCGGCGCCTCGAAGGCGCCTCGGACCTCATGGCGCTCCGCCACCACCG 3831
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACCTCGCCGTACGCGGCGCCTCGAAGGCGGCCTCGGACCTCATGGCGCTCCGCCACCACCG 480

Qy  3832 CACCCACGGCCTTGGACGTCCGGGTGACCCGCTGTTCGAACAACTTCGGCCCACCAGCA 3891
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CACCCACCGCCTTGGACGTCCGGGTGACCCGCGTCGTTCGAACAACTACGGCCCACCAGTT 540

Qy  3892 TCCCG 3896
         || ||
Db   541 CCCGG 545
```

RESULT 7
US-09-860-846-13
; Sequence 13, Application US/09960846
Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-13

```
Query Match             6.5%;  Score 447;  DB 9;  Length 1014;
Best Local Similarity  56.8%;  Pred. No. 3.4e-82;
Matches  655;  Conservative   0;  Mismatches 320;  Indels  6;  Gaps  1;

Qy  3345 TCCTCGTCACCGGCGGAGCGGGCTTCATTCGGTCCGCTACGTCCGGCTCCGTGTCGC 3404
         | || ||||||| ||| |||||||||||| | |||| ||||| || || |||| |||||
Db     8 TTCTGGTGACCGGAGGTGCGGGCTTCATCGCCGTCACTTCGTCGCGCCAGCTCCTCGCCG 67
```

```
Qy  4752 CGAGTGCTGCACGAGCTCCTCGGACTCCCGCCGCGCCGACACTCCCAGGTGCTCCCCCACGCC 4811
         || ||| |||| ||||||||||  ||  ||||  |  |||  |    |||  |  || ||
Db   7537 GCCGACATGACGGCCCTCCGCGAGCTCGCGCGACCGGCCACGGCCTGCACATCGTCGAGGAC 7478

Qy  4812 GGGCCGGAACCCCTCGATCTTCCGCTGCACCGGCGCACGAAGCGGGCGGGCCCGACGGCGAG 4871
         ||| | |  ||  ||| |||  |||  |  ||  |     || ||| |||  || |  ||||
Db   7477 GCCGCGGCAGGCCCACG---GCGCCCGCTACCGGGCGGCCGGATCGGCCGGCCGGGTCGTCG 7421

Qy  4872 GCCGACGGCGCTCGTGCCCCAGCTCCGCCAGTGCCCCAGGCGCCAGGCGCATCGCGGCCCG 4931
         ||| |  ||| |  || |||||  ||  ||  |    ||  |  ||  |    ||  |||||
Db   7420 GTGGCCGCGTTCAGCTTCTACCCGGGCAAGAACCTCGGCTGCTTCGGCGACGGCGCGGCC 7361

Qy  4932 GTGCTGCCGGGGAGCAGGAAGGGTGCCTGCGATCCGGGGCGGAGAACCCGTTCACGAGGACGAA 4991
         |||| |  |||  ||  |  |  | ||  ||  |  ||  |  ||| |  ||  |  |||  |
Db   7360 GTCGTCACCGGCGACCCCGAGCTGCGCGACGCTGCCGATGCGTCCGCAACTACGGCTCG 7301

Qy  4992 GGGCA 4996
         ||| |
Db   7300 CGGCA 7296
```

RESULT 6
US-09-922-683-3
; Sequence 3, Application US/09922683
Publication No. US20020192793A1
; GENERAL INFORMATION:
; APPLICANT: DECKER, Heinrich
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
GLA.O AND THEIR USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/922,683
; FILING DATE: 07-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/194,905
; FILING DATE: 1999-12-01
; APPLICATION NUMBER: DE 19622783.6
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granados, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 026083/0193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO:3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-922-683-3

```
Query Match             7.6%;  Score 522.6;  DB 9;  Length 546;
Best Local Similarity  97.4%;  Pred. No. 1.6e-97;
Matches  531;  Conservative   0;  Mismatches 14;  Indels  0;  Gaps  0;
```

Query Match          6.5%;   Score 447;   DB 9;   Length 1014;
Best Local Similarity  66.8%;   Pred. No. 3.4e-82;
Matches  655;   Conservative    0;   Mismatches  320;   Indels    6;   Gaps   1;

Qy 3345 TCCTCGTCACCGCGGAGCGGGCTTCATTCGTCGCCTACGTCCGCCGGCTTCCGTGTCGC 3404
             |   |  |||| ||||| ||||| ||| | |||||| |   || | |||||||||
Db    8 TTCTGGTGACCGGAGGTGCGGGCTTCATCGGCTTCGTCGGCAGCTTCCTCGCCCG 67

Qy 3405 CCGGGGCCCCCGCGGCGGTCGC------GGTGACCGTTCCTCGACAAACTCACCTACGCCG 3458
             |  | ||||   | |  ||    ||| |||| || ||| || ||| |||||||||||
Db   68 GGGCGTACCCCGACGTGCCCCGACGGGCGACCCGCGACTGCGTTCGTCCCACGGCG 127

Qy 3459 GCAGCCTCGCCCGCTGCGACACGCGTCGACACGCTGGCCGGCGACGACATCGTGCACT 3518
            || |  |||| |   || || || | ||| || ||||| |  |  |  ||||||| |||
Db  128 GCAACCCGGCCAACCTCGCCCGCTGGACGCGCGACCCGGCGTTCGTCCACGGCG 187

Qy 3519 ACGTGTGGCACACCGCGCTCGTCGCAGCCTCCTGACCAGCCTGCTGCACGACATCGTGCACT 3578
            || |    |  || |   ||  |||| | |  |||| |||||||||||||||  || ||||
Db  188 ACATCCCGGCGACGCGGCCGGCTCCTCGCCCGGCGTGGAACTGCGGCCGGCGTGGAACTGCGGCCATCGTCCACT 247

Qy 3579 TCGGGGCCGAGTCGCACGTCGGCTGCCTTCACCCCGCACCA 3638
            |||| |  || |||| |    |  ||  ||||  |||||||||||
Db  248 TCGCGGCCGAGAGCTCCTCGCCCGGCGGGCGCGCGGGCGTCGGGGTCG 307

Qy 3639 TCGCGGCCGAGTCGCACGTCGCCATCACCCGCGCTCCATCACCCGCGCCACCA 3698
            |||||||||||| |||||| | | ||| ||||||| ||  ||| || | || |||
Db  308 TCGGCGCCGAGAGCACGACGTGGACCGCGCCGTCGCGCCGCGTCGGGGTCG 367

Qy 3699 ACGTGCTGGCACACCGCCTCGTCGGCGCGGCCGCCGCGGGAGCGGGGGAGAGCG 3758
            ||||||||||||||||| ||||| |||| |||| | |||  || |  |   || |||
Db  368 TGCACGTCTCCACCAGGCGTTCCGACCTCGACGTCGCGCGGCGGGCGCGCGAGACCA 427

Qy 3759 TCGGGGCCGAGTCGGCGCGCTGCGACGTCCGGGGGGAGAGATCTACAAACATCG 3818
            ||||| | | || |||  |||  ||| ||| | ||  | |||  |   | |||| || |
Db  428 GCCCGCTGGAGCCCAACTCCGGCGTCGCCCGGCGTCCGAGCCTCGACCTGTTGCCC 487

Qy 3819 TCGCCCACCACCCGCACCCGCTGTCGGGTGACCCGTGTTCGAACAACTTCG 3878
            ||| | |||| || || | |  ||| || ||||||  | |||| || ||||
Db  488 GCGGCCTACCACCGGACGCGCTCGCCGCTACCGCGCATCACCCTGGACTCGCTGGACTACG 547

Qy 3879 TCGGGGCCCCACCAGCATCCCGAGAAGCTCATACCGCGCTTCCTGACCAGCCTCCTGTCCCGGCG 3938
            |||||||||||   ||||| |||  ||||| |    |||  | | |  |||||||||  || |  |||
Db  548 TCGCACGTCTCCACCAGGCGTTCCGACCTCGACGTGGGTCGATCGACTCCGGCGACTCCTGGCGTCGG 607

Qy 3939 GCAGCGGTTCCCCTCTACGGCGCGGCGCACGGGCGCACGTCGCACGTCGACGACC 3998
            ||| |||||  | |  ||||||| ||||| ||| ||  |||||||||||||||||
Db  608 GGAGCGTCCGCGTCGACGACGCGCGAACGTCCGGAGTGGGTGCACACCGACGACC 667

Qy 3999 ACGTCAGGGCCGTCGAACTCGTCGGCGCCGGGAGAGATCTACAAACATCG 4058
            | |   | |  ||||||| |  |||||||||| || || ||   | || || ||
Db  668 ACTGCCGGGCCACCTCGCTGCCCAACTGGAGCTCACGACGAGCTCACGGCGGCG 727

Qy 4059 GGGGCGGGCACCTGCTCAGGCGTCTCACGCGACGCTGCGGCG 4118
            |||| | |||  ||||| | |  || | | | ||  |  |||||
Db  728 GGGCGGGGCGACGTGAGCTGACGAGTGTACGGGTGCCGAGAGCA 787

Qy 4119 CGGGCCCGGAGCGCCGCCATCGTCCAGGTGCGAGAACCGCAAGGGCGTACGCGG 4178
            |  |||| |||| ||| |  |||||  |   | |||  | ||||| ||  | ||||
Db  788 CCGACTGGTCTCCGGCGCCCGGAGGTGCGCCAGGGCGTCGGCGGGTGCGCCACCA 847

Qy 4179 TCGACCACAGCAAGATCACCGCCGAACTCGGTTACCCGGCCGACCTTCGCGACCG 4238
            |||| | ||||| || ||||  || ||| || | |||||| || ||  ||||||| |
Db  848 TCGACGGGCGGACGATCGAGCGCGGACGTCGGCTACCCCGCGTCGCTCCTTCGCGGACG 907

Qy 4239 CGCTGCGCGACACCCGCAAGTGGTACGAGCGGCACGAGGACCTGTGGGCGTCCCCTGCTCG 4298
            | || | || || ||  |  | |||  |  | ||| | |   ||  | |  |||| || |
Db  908 GCCTCGGCGGAGCTGGACCAACCCGGTCCGGCTCCGGGGAGAACCGCGGGTCGCGTCAAGG 967

Qy 4299 CGGCGACATGACGTCGGGCGCCG 4319
            | | |  |  || | ||| ||||||
Db  968 CGACCGCCCCCAGCTGCCCG 988

RESULT 8
US-09-988-384B-13
; Sequence 13, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.

```
Db    788  CCGACTGGTCTCGGTCCGGAAGGTCGCCGCCGCCAAGGGCCACGACCTTCGCGCTACTCCC  847
Qy   4179  TCGACCACAGCAAGATCACCGCGGAACTCGGTTACCGGCCGCGCGACTTCGCGACCG     4238
Db    848  TCGACGGCGGCGAGATCGAGCGCGAGGTCGCGAGTCTCCTTCGCGGACG              907
Qy   4239  CGCTGGCCGCGACACCGGCGAAGTGGTACGACGCGCACGAGGACTGGTGCGGCGTCCCCGCTCG  4298
Db    908  GCCTCGCGCGGCGGGACCGCTGGTACCGGGAGAACCGCGGCCTGTGGGAGCCGCCGCTCAAGG   967
Qy   4299  CGGCGACATGACGTCGGGCCG  4319
Db    968  CGACCGCCCCGCCAGCTGCCCG  988
```

RESULT 9
US-09-836-821-13
; Sequence 13, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-13

Query Match       6.5%;   Score 447;  DB 9;  Length 1014;
Best Local Similarity   66.8%;  Pred. No. 3.4e-82;
Matches  655;  Conservative   0;  Mismatches  320;  Indels   6;  Gaps    1;

```
Qy   3345  TCCTCGTCACCGGCGGAGCGGGCTTCATTCGCTCGCCGCTCCGCTCCTCGTCGC  3404
Db      8  TTCTGGTGACCGGAGGTGCGGGCTTCATCGCGTCGCCGCAGCTCCTCGGCCG     67
Qy   3405  CCGGGGCCCCGGCGGGCGTCGC----GGTGACCGTCTCCGACAAACTCACCTACGCCG  3458
Db     68  GGGCGTACCCGACGTGCCCGCCCGATGAGGTGATCGTCTCCGACAGCGCCTACACGCGG  127
Qy   3459  GCAGCCTCGCCGGCGGCCTGCACCGGCGCGGTGACCATCCCGGCCTCACCTTCGTCCAGGGCG  3518
Db    128  GCAACCGCGGCCAACCTCGCCCGGTGGACCGGCGGACCCGCCGTTCGTCGTCCCACCGGCG    187
Qy   3519  ACGTGTGCGACACCGGCGCGTCGTCGACACCGGCGCGGCGCACGACGCATCGTGCACT  3578
Db    188  ACATCCGCGACGACGCCGGCCTCTCGCCCGGGAAACTGCGCGGCGCATCGTTCCACT    247
Qy   3579  TCGCGGGCCGAGTCGCACGTCGCACGTCGACCGGCGCCTTCACCCGCCACCA  3638
Db    248  TCGCGCCGGAGAGCCAGACCGGTGACCGCGTGTCCATCACCGGCGCGTCCTCCGTTCACCGAGACCA  307
Qy   3639  ACGTGCTGGGCACCCAGGTCCTGCTCGACGCCGCCGCCACGGTGTGCGCACCTTCG  3698
Db    308  ACGTGCAGGGCACCGCACCGCGAGACCGCTGCACGGCGCCGGGGTCGGGCGGGTCG  367
Qy   3699  TGCACGTCGTCCACCGACCGAGAGGTGTACGGCGTACGGGTTACGGCGCTCCGGGCGGAGAGCG  3758
Db    368  TGCACGTCTCCACCGACGAGGTTCACGGCTCGATCGACTTCCGGCTCCTGGACCGAGAGCA  427
Qy   3759  ACCCCCTGTTCCGACCTTCGGACCTGCGGCGTACGCGGGCGGCTCGAAGGGGCCTCATGGCGC  3818
Db     68  GGGCGTACCCGACGTGCCCGCCGATGAGGTGATCGTCTCCGACAGCGCCTCACCTACGCGG  127
```

```
Db    428  GCCCGCTGGAGCCCAACTCGCCCTACGCCGGCGGTCCGAGGCCGGCCGTCCGACCTCGTTGCCC  487
Qy   3819  TCGCCCACACCGCACCGAGCCGGCTGGACGTCCGGGTGACCGTCGTTCGAACAACTCG  3878
Db    488  GGCGCCTACCACCGGACGTACGGCCTCGACGTACGGATCACCCGCTGCTGCAACAACTACG  547
Qy   3879  GCCCCCACCAGCATCCGAGAAGCTCATACCCGCGTTCCTGACCAGCCTCGCTGTCCCGGCG  3938
Db    548  GGCCGTACCAGCACCCGAGGCGGAGCTCATCCCCCCTTCGTGACGAACCTCCTCGACGGACG  607
Qy   3939  GCACCGTCCCCTCCTACGGCGACGGGGCGCACGTGGCCGACTGGCTGCACGTCGACGACC  3998
Db    608  GGACGCGTCCCGCTGTACGGCGACGGCGCGGCGAACGTCCCGCCGAGTGGGTGCACACCGACGACC  667
Qy   3999  AGTCAGGGCCGTCGAACTCGTCTCCGGCCGTGTCGGGCCGCCGGGGGCCGGGAGAGATCTACAAACATCG  4058
Db    668  ACTGCGGGGCATCGGCCCTCGTCCTCGTCCGGGGGGCCGGGAGATCTACCACATCG  727
Qy   4059  GGGGCGGCACCTCGCTGCCCAACCTGAGCTCACGCACCGGTTGCTGCACGTGCGGCGG  4118
Db    728  GCGGCGGCTGGAGCTGACCAACCGCAGCTCACCGCATCCTCCTGGACTGCTCGGCG  787
Qy   4119  CGGGCGCGGAGCGGCCATCGTCCACGTCGAGAACCGGCAAGGGGCACGACCGGCCGCTACGCGG  4178
Db    788  CCGACTGGTCTCGGTCCGGAAGGTCGCCGCCGCCAAGGGCCACGACCTTCGCGCTACTCCC  847
Qy   4179  TCGACCACAGCAAGATCACCGCGGAACTCGGTTACCGGCCGCGCGACTTCGCGACCG  4238
Db    848  TCGACGGCGGCGAGATCGAGCGCGAGGTCGCGAGTCTCCTTCGCGGACG  907
Qy   4239  CGCTGGCCGCGACACCGGCGAAGTGGTACGACGCGCACGAGGACTGGTGCGGCGTCCCCGCTCG  4298
Db    908  GCCTCGCGCGGCGGGACCGCTGGTACCGGGAGAACCGCGGCCTGTGGGAGCCGCCGCTCAAGG  967
Qy   4299  CGGCGACATGACGTCGGGCCG  4319
Db    968  CGACCGCCCCGCCAGCTGCCCG  988
```

RESULT 10
US-09-861-289-13
; Sequence 13, Application US/09861289
; Patent No. US2002011089A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-13

Query Match       6.5%;   Score 447;  DB 10;  Length 1014;
Best Local Similarity   66.8%;  Pred. No. 3.4e-82;
Matches  655;  Conservative   0;  Mismatches  320;  Indels   6;  Gaps    1;

```
Qy   3345  TCCTCGTCACCGGCGGAGCGGGCTTCATTCGCTCGCCGCTCCGCTCCTCGTCGC  3404
Db      8  TTCTGGTGACCGGAGGTGCGGGCTTCATCGCGTCGCCGCAGCTCCTCGGCCG     67
Qy   3405  CCGGGGCCCCGGCGGGCGTCGC----GGTGACCGTCTCCGACAAACTCACCTACGCCG  3458
Db     68  GGGCGTACCCGACGTGCCCGCCCGATGAGGTGATCGTCTCCGACAGCGCCTCACCTACGCGG  127
```

```
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/922,683
    FILING DATE: 07-Aug-2001
    CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 09/194,905
    FILING DATE: 1999-12-01
    APPLICATION NUMBER: DE 19622783.6
    FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
    NAME: Granados, Patricia D.
    REGISTRATION NUMBER: 33,683
    REFERENCE/DOCKET NUMBER: 026083/0193
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (202) 672-5300
    TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 4:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 541 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-922-683-4
```

```
Query Match        4.0%;     Score 276.2; DB 9;   Length 541;
Best Local Similarity 69.6%;  Pred. No. 1.9e-47;
Matches 374;  Conservative 0;  Mismatches 163;  Indels 0;  Gaps 0;

Qy  3362 GCGGGCTTCATTCGCTCCGCCTACGTCCGCCGGGCCCCGGGCGGCCCCGGCGGGC 3421
              |||||||||| ||| |||||||  |||||| |||||||||  ||| ||| |
Db   537 GCGGGGTTCATGGGCTCGCACTACGTCCCACCCTCCTCGGTCCGGACGGGCCGCCCGGAC 478

Qy  3422 GTCCGGTGACCGTCCTCGACAAACTCACCTACGCCGCCTCGCCGCCTCGACGCGG 3481
              | ||| ||||||| ||||||  ||||| ||||||  ||||||  |||||||| |
Db   477 GCCGTCGTCACCGTGCTCGACGCCGTCGACGCTACGCCGGCAACCTCGCCAACCTCGACCCG 418

Qy  3482 GTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTCGCACACCGGCTCGTC 3541
              || |||||||  |||||||| ||||||||||  ||| |||  ||||| ||| ||| |||
Db   417 GTCCCGACCACCCGGCCTTCGCCGCCTTCGTCGTCCAGCATCGTGCGACGCGGGACCTGGTC 358

Qy  3542 GACACGCTGGCCGCGCGGCGACGACGACATCGTGCACTTCGCACGTCGCACGTCGAC 3601
              ||| ||  ||||| ||||  |||| ||||||||||| |||||||||||| |||||||
Db   357 GACCGGGTCATGGCCCGCCAGGACCGCCAGGGTCGTGCACCTCGCCGCCGAGTCGCACGTGGAC 298

Qy  3602 CGCTCCATCACCGACCACGGTGCCTCACCCGCCACCAACGCTGCGGCCAGTGCTCCTG 3661
              ||||| ||| |||||  |||| |||| | |||||  ||||||| |||  |||| |
Db   297 CGCTCGTCGCTCGATGCCCGCGGTTCGTGCGGACCAAGCGCAGCAGACACTC 238

Qy  3662 CTCGACGCCCGCGCTCCGACGCCCACGGTGTGCGCCACCTTCGTGCACGTCGCCGACGAGGTG 3721
              |||||||| |||||  || ||||| ||| | ||||||||||||||| |||| |||||||||
Db   237 CTCGACGCCGGCGCCCTCCGGGACGCGGCCTGGGCCCCCTTCGTGCAGGTTCCACGACGAGGTC 178

Qy  3722 TACGGCTCCCTCCCCGACGGGGCCGCCGGGAGAGCGACCCCGTTCTCCGACCTCGCCG 3781
              ||||||||  | ||| |||| | || ||||||| ||||| |||| | |||||  |||
Db   177 TACGGCTCGTTGGAGACGGGGTCGTGGACGGAGGACGAGCCGCGCTGCGCCCAACAGCCCC 118

Qy  3782 TACGCCGGCGTGAAGGCGGCCTCGAACGCCTCATGGCGGTCCGCCACCACCGACCACCGCG 3841
              || ||| || |||||||| ||  | ||  ||||||||  |||| ||||| | | |  |||
Db   117 TATGCCACCTCCAAGGCGTCGGGCGACCTGCCTCGCGGCTGCGCCGGCATGCAGCGTCAGCCAGGGC 58
```

```
Qy  3459 GCAGCCTCGCCCGCCGCTGCGCGGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCG 3518
              ||| ||||| ||| |||   || |   ||   || |||| |||  ||||| |||||||| |||
Db   128 GCAACCGCCGCCAACCTCGCCCGGTGACGGGGACCCCGCCGCTTCGTTCCACGGCCG 187

Qy  3519 ACGTGTCGCGACACCGGCCTCGTCGCACACGCTGGCCGCGCGGCGCACGACATCGTGCACT 3578
              || || |||  |||| ||||| ||   ||  || ||| ||| || || |||||||| |||||
Db   188 ACATCCCGGACGCGACGGCCTCGCTCCGGGGAACTGCCGCCGTGGGCCATCGTCCACT 247

Qy  3579 TCGCGGCCGAGTCGCACGTCGACCGTCCATCCATCACCGACACAGCCGGTGCCTTCACCCGCCACCA 3638
              ||| ||||| ||||| | ||  || |  |  ||||| || |||    ||| |   ||||
Db   248 TCGCGGCCGAGAGCCACGTGGACCGTCCATCGGCGCGCTCCGTTCACCGAGACCA 307

Qy  3639 ACGTGCTGGCACCAGTCCTGTCTGCGACGCCCGCGCTCCGCCACCGTGCGCACCTTCG 3698
              ||| |||||||||||| || || ||| |   ||| ||||  || |||||| || |||||
Db   308 ACGTGCAGGGCACGCAGACGCGTCCAGTTCCAGGGCGTCGACCCGCGGTCGG 367

Qy  3699 TGCACGTCTCCACCGACGAGGTGTACGGCTCCCTCCCCCGCACGGGCCCGCGCCGGGAGAGCG 3758
              |||||||| |||| | ||   || ||||||| |||||| |  |||| ||| ||| ||||| ||
Db   368 TGCACGTTCCACCGACGAGGTGTACGGCTCCCTCCGGATCCTCGGCACCGAGACCCGCCAGAGCA 427

Qy  3759 ACCCCGTGCTTCCGACCTCCGCCCGTACGCGGGCCGTCGAAGGCGGCCTTCATGGCGAC 3818
              ||||| | || ||||||||||| |||||| ||||| | || ||  || || | ||| ||
Db   428 GCCCGCTGAGCCCAACTCGCCCTACGGCGGTCCAAGGCCGCTCCGACCTCGTTGCCC 487

Qy  3819 TCGCCACCGACCACCCGGCCTCGGGTGACGTCCGGGTGACCCGTGTTCGAACAAGTTCG 3878
              |||  ||| ||| ||  | || | |||  || ||   || || | |  |  ||| |||||
Db   488 CGGCTACCACCCGACGACGTACGGGCCTACGGACGCTAGGCGCTCGTGCAACAACTACG 547

Qy  3879 GCCCCCCACCAGCATCCGAGAAAGTCATACCGGCGCTTCCTGACCAGCCTCGTCCTGCCGGCG 3938
              || || | |||| |||  || |||| |   ||| |  |||| |  ||||| ||||  |||||
Db   548 GGCGTACCAGCACACCCCGAGAAGCTCATCCCCCTCTTCTCGTGACGAACCTCGACGGCG 607

Qy  3939 GCACCGTTCCCTCTCACGGCGACGACCGTGCGCGACGTGCTGGTCGACGTGGACGACC 3998
              || ||| | | || ||| | ||| | |  |||| |||| ||| ||  ||||| |||||
Db   608 GGACGGTCCCGGCTCGTGTACGGGCGACGCGGCGGGAACGTCCCGGGAGTCGGGCAGCAGCAAC 667

Qy  3999 ACGTCAGGGGCCGTCGAACTCGTCCCGGTGTCGGCCGTGTCGGGCGGCCGGCGGCCGGCG 4058
              ||| || ||  | || |  ||||||  ||| | ||||  || || || || || |||| |
Db   668 ACTGCCGGGGCCATCGCCGCTCGCTCCGGCGGCGGCTGGGCCCCCTTCGTGCAGGTTCCACCATCG 727

Qy  4059 GGGGGCGCCACCTGCGCCAACCTGGAGCTCACGGCACCGCCGGCGCCGGCACCGTGCGGCG 4118
              |||||| || ||| ||||  |  || ||||| || |  |  |||| | || |||||
Db   728 GCGGGCGGCCGTTGGAGCTGACCAACCGGCAGAACTCACCACCGTCGCATCCTCCGGACTCGCTCGGCG 787

Qy  4119 CGGGCCCGGAGAGCGCATCGTCCACGTCGAGAACCGGCAGAACCGGCGCGTCTACGCGGG 4178
              |||| ||||||| || ||||   |  ||||| | ||| ||||| ||  |  || |||| ||
Db   788 CCGACTGGTCTCCTCGGTCCGGAAGGTGCCGCGAAGGTGCAGGCCACGAACCTGCGCTACTCCC 847

Qy  4179 TCGACCACGAAGATCACCCGCGGGACTCGGTTACCGCGCCGGCGGCGACCGACTTCGCGACCG 4238
              ||| || ||  |||| || || ||| ||||  ||| |  || ||| ||| |||  ||||
Db   848 TCGACGGCGGGCAGATCGAGCGCGGACGGCGGAGCTCCGGCGCTACCGCCGCCAGGTCTCCTTCGCCGACG 907

Qy  4239 CGCTGGCGACACCGCGGAAGTTGTACGAGCGGCACGGCGAGTGGTCGCGTCCCCGCTCG 4298
              |||| |||  ||||  |  || | | || ||| | ||| | |||  || || |  || ||
Db   908 GCCTCGCGCGACGTCGTCGCTGGTACCGGGAGAACGCGGCCGGGCGGGGAGCGCGCTCAAGG 967

Qy  4299 CCGGGACATGACGTCGGGCCG 4319
              | || || ||  |||| ||
Db   968 CGACCGCCCCGCAGCTGCCCG 988
```

```
RESULT 11
US-09-922-683-4/c
; Sequence 4, Application US/09922683
; Publication No. US20020192793A1
GENERAL INFORMATION:
    APPLICANT: DECKER, Heinrich
    TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
        PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES-GLAUCESCENS
        GLA.O AND THEIR USE
    NUMBER OF SEQUENCES: 13
    CORRESPONDENCE ADDRESS:
```

```
Qy   3842 CTGGACGTCCGGGTGACCCGTGTCGAACAACTTCGGCCCCCCACCAGCATCCCGAG 3898
          ||||||| || | ||||| | |||||||| |||| ||||||| |||||||| ||.||
Db     57 CTGGACGTGCGCATCACCCGCTGCTCCAACAACTACGGCCCCTACCAGCACCCGGGG 1

RESULT 12
US-09-738-626-373
; Sequence 373, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT: MIZOGUCHI, HIROSHI
;  APPLICANT: ANDO, SEIKO
;  APPLICANT: HAYASHI, MIKIRO
;  APPLICANT: OCHIAI, KEIKO
;  APPLICANT: YOKOI, HARUHIKO
;  APPLICANT: TATEISHI, NAOKO
;  APPLICANT: SENOH, AKIHIRO
;  APPLICANT: IKEDA, MASATO
;  APPLICANT: OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 373
;  LENGTH: 1131
;  TYPE: DNA
;  ORGANISM: Corynebacterium glutamicum
US-09-738-626-373

Query Match       3.0%;    Score 205.2; DB 9; Length 1131;
Best Local Similarity   56.1%;   Pred. No. 4.5e-33;
Matches  497;  Conservative    0;  Mismatches 368;  Indels  21;  Gaps  5;

Qy   3338 ACGACCATCCTCGTCACCGGAGCGGGGCTTCATTCGCTCCGCCTACGTCCGCCGGCTC 3397
          || ||| |||||| ||||||| |||||| |   ||| || || |||||||||||||
Db    118 ACTTCTTTGCTTGTGACCGGAGGTGCCGATTTATCGGCGCCAACTTCGTCCGCCAAACC 177

Qy   3398 CTGTCGCCCCGGGGCCCCCCGGCGGTCGCGGTGACCGTCTCTCGACAAACTCACCTACGCC 3457
          |   ||       ||| | |   |   |||| |   || | ||| ||||||||||||||
Db    178 GTAGAGC---AGCACCCTGAATACACCCCACATCACGGTGCTGGATAAACTCACCTACGCA 234

Qy   3458 GGCAGCCTGCCCGCTGCCCGACGCGGTGCGTGGCCATCCCGGCTCACCTTCGTCCCAGGGC 3517
          || |  | |  ||| ||  || ||| |   |||   | | ||| || ||| || || |||
Db    235 GGAAACGCCGACAATCTCAAAGGCCTCCCCGACCAGCAAAG---TAACCCTCACCGAAGGC 291

Qy   3518 GACGTGTGGCGACACCGGCGTCGTCGACGACGCGCTGGCGCGCGGCACGACGACATCGTGCAC 3577
          | ||| || |||| |||| |   ||||||| || ||   |  ||| || || || ||||
Db    292 GATATCTGCGATGCTGAATTAGTCGACTCCCTGTCAAAGACCACGACGACAATCACAGTCCAC 351

Qy   3578 TTCGCGCCCAGTCGCACGTCGACCGCTCCATCACCCCGCACCGGTGCCTTCACCCGCCACC 3637
          |||||| || ||| ||| | || |   || |  | ||| || |||| || |||||||
Db    352 TTCGCAGCAGAATCCCACACGACAACTCCCTCAACGACCCCTCCCGTTTGTTCACACT 411

Qy   3638 AACGTGCTGGGCACCCAGGTCCTGCTCGACGCGCCGCTCCCCACGGTGTCGCCACCTTC 3697
          || | | | || || |||| | |||   | | ||  ||  |||||| ||  | |||
Db    412 AACCTCATCGGCACCTTTGCTCCTCGTAGAAGCAGTCAGTCCAAGCACACAAACG--CTTC 468

Qy   3698 GTGACGTCTTCCACCGACGAGGTGTACGGCTCCCTCCCCGCACGGG-------CCGCC 3748
          ||  ||||| | |||||||      |||| ||  | | |  | |||      |||||
Db    469 CACCACCATTCCACCGATGAAGTCTTCGGCGATCTAGAGTCGGATGATCCAAACCGCTTC 528

Qy   3749 GCGGAGAGGCGACCCCCCGTGCTTCCGCACCTCGCCTACGCCGGCGCCTCGAAGGCCTCGGAC 3808
          | |||||   |   || | |||| ||||  || | ||||||||| ||||| ||
Db    529 ACTGAAACCACCGCCTACAAGCCATCGTCTCCATATTCTGCAACCAAGCCAGTACGGGTCTGAT 588
```

Right column:

```
Qy   3809 CTCATGGCGCTGCCCCACCACCGCACCCACGGCCTCGGACGTTCCGGGTGACCCGCTGTTCG 3868
          |.||| | | || |||| ||||| |||||||| || | |
Db    589 CACTTGGTACACGCATGGATCCGCTCCTTCGGAATCCAGGCAACCATGTCTAACTGCTCC 648

Qy   3869 AACAACTTCGGCCCCACCAGCATCCCGAGAAGCTCATACCGCCGTTCCTGACCAGGCTC 3928
          |||| | |||| ||  || ||||  || ||||||| || | ||||||||| |
Db    649 AACAATTACGGTCCCTACCAGCACATTGAAAAGTTCATCCCCCGCAGATCACCAATATT 708

Qy   3929 CTGTCCGGCGGCACGTTCCCCTACGGCGACGGCGCAGCTGCCGACTGGCTGCAC 3988
          || ||  | ||| ||||  |   ||||||| || |||||| |||||| ||||||
Db    709 CTGGCCCGCCTGACACCAAAACTTATGGAACCGGCAGTCCGGACTGGATCCAC 768

Qy   3989 GTCGACGACCACGTCAGGCGCTCGAACTCGTCGTCCGGTGTCGGCGCCGGCGGAGAGATC 4048
          |||| | || | |   |  ||   |   |   || ||||||| || |||||
Db    769 GTCGATGATCACAATGACGCGCCTCCACCTGATCCTGAGTAAGGCAAGATCGGCGAAACC 828

Qy   4049 TACAACATCGGGGGCGGGCACCTCGC---TGCCCAACCTGGAGCTCACGCACCGTTGCTC 4105
          ||| || ||||||| || || | | |    || |  | ||||| || |||   | |
Db    829 TACATCATCGGCGCCGGACAACATCATGTGTGAATAACAGCAGGTCATCGAGCTTATTGT 888

Qy   4106 GCACTGTGCGCGGCGCGGCGGCCGCATCGTCCACGTCGAGAACCGCAAGGGGCACGAC 4165
          ||| ||||  ||  | ||| |  ||||||||  ||||  || |||| |||||| |||
Db    889 GAACTCATGGGCCTCGACAAAAACGCATACGAGCACGTCGCAGACCCGCCGCCACGAT 948

Qy   4166 CGGCGCTACGCGGTCGACCAGCAAGATCACCGCCGGAACTCGGTT 4211
          |||  |||| |   || |  | | |  |  || || ||  |||||
Db    949 ATGCGTTACGCCATGATTCCACCAGCTGCCACCGAGCTCCGGCT 994

RESULT 13
US-09-738-626-536/c
; Sequence 536, Application US/09738626
; Publication No. US2002019705A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT: MIZOGUCHI, HIROSHI
;  APPLICANT: ANDO, SEIKO
;  APPLICANT: HAYASHI, MIKIRO
;  APPLICANT: OCHIAI, KEIKO
;  APPLICANT: YOKOI, HARUHIKO
;  APPLICANT: TATEISHI, NAOKO
;  APPLICANT: SENOH, AKIHIRO
;  APPLICANT: IKEDA, MASATO
;  APPLICANT: OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 536
;  LENGTH: 1344
;  TYPE: DNA
;  ORGANISM: Corynebacterium glutamicum
US-09-738-626-536

Query Match       2.4%;    Score 163.6; DB 9; Length 1344;
Best Local Similarity   49.7%;   Pred. No. 1.3e-24;
Matches  601;  Conservative    0;  Mismatches 579;  Indels  30;  Gaps  6;

Qy    720 CGCGATCACCCGCTCGACGGCCGTCTCCAGCAGGCCGGCCCTCGTCGAGCAGCGCGCTC 779
          || | |  |||||  ||||  |  ||| ||||||| || ||  ||||||||| ||||||
Db   1329 CGCGGTTTCGCGCTCTAGGGCGTGCGACTAACACCTGAAGACCATCACCGAGAGTGTCCTC 1270

Qy    780 GTCGAGGGTGAACGGCCGTGCCAGCCCGGCGGATGTGGC---CGCCCAGGAGGTGCCGCAG 836
          || | | || |  | |     || ||   || | |||||    |.|  || |||| ||||
Db    780 ...
```

us-09-922-683-7.rnpb

Thu May 29 15:11:30 2003

```
Qy    1905 CTCGGGGAAG 1914
            ||||||||||
Db     144 GTCCACCACG 135
```

RESULT 14

US-09-860-846-11

; Sequence 11, Application US/09860846
; Patent No. US2002016474 2A1
; GENERAL INFORMATION:
;   APPLICANT: Sherman, D.H.
;   APPLICANT: Liu, H.
;   APPLICANT: Xue, Y.
;   APPLICANT: Zhao, L.
;   TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;   FILE REFERENCE: 600.438US1
;   CURRENT APPLICATION NUMBER: US/09/860,846
;   CURRENT FILING DATE: 2001-05-18
;   PRIOR APPLICATION NUMBER: 09/105,537
;   PRIOR FILING DATE: 1998-06-26
;   NUMBER OF SEQ ID NOS: 43
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
;   LENGTH: 879
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-11

Query Match             2.3%;     Score 157;  DB 9;   Length 879;
Best Local Similarity  53.6%;     Pred. No. 2.9e-23;
Matches  397;  Conservative  0;  Mismatches  335;  Indels  9;  Gaps  3;
```

```
Qy    2269 TGAAGGCCCTGGTCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTCACCCACACCG 2328
            |||||                  ||||| | |||||||||||  ||||||||||||||||
Db       2 TGAAGGGATAGTCTGGCGGCGGAGCGGGAACTCGGCTGCATCCGGCGACCTCGGTCA     61

Qy    2329 CCGCCAAGCAGCTGCTCCCCATCGCCGGCGGCCAACAAGCCCGTGCTCTTCTACGCGGCTG 2388
            |||||            || ||| |||  ||||||||||||||| ||| |||   || |||
Db      62 TTTCGAAGCAGATTCTTCCGGTCTACAACAAACCGATGATCATCTACTACCGCTGTCGGTTC 121

Qy    2389 TCGGCCGCGGGGGGTGTCCGGGAGGCCGGCGTCGTCGTGGCGGCGTACGCCGGGGAGATCC 2448
            |||                  ||| || ||||  |||   ||  |||  |||| || ||
Db     122 TCATGCTCGGCGGTATTCGCAGAGATTCAAATCATTCGGTTACGGCTTACGACATCGAACTCT 181

Qy    2449 GCGA---ACTCACCGGCGACGGCGACCACCGCGTTCGGTTACGCGGTTACCATCACCTCCACCAGC 2505
            ||                  ||||||| |||  |||| || |||  ||||| ||| ||||||
Db     182 TCCAGTCGCTTCTCGGAAACGGCAGCAGGCACCTGGAATAGAACTCGACTATGCGGTCCAGA 241

Qy    2506 CCCGCCCGCTCGGTCTCGGCGTCGGTGCCATCGCCGGGCGACGACG 2565
            ||                  || ||||| | || ||||||||||| |||||||||
Db     242 AAGAGCCCCGCAGGAATCGCGGCGACGCACTTCTCGGCGGAGCGGAGCGACGAGCACATCGGCGACGACA 301

Qy    2566 ACTTCCTGCGTGTACCTGGGGGACAAACTACCTGCCC--CAGGGCGTCACCGGACTTCGCCC 2622
            ||                  ||  || ||||||  ||||||| ||||||
Db     302 CCTGCGCCCTGATCCTGGGCGACACACATCTTCCACGCGGGCCGGCCTCTCACACGCTCCTGC 361

Qy    2623 GCCAATCGGCGCCGATCCGCGCGCCGATCCGCGGCGCTGTCGGTCTGCGTCGGCGGACCCGT 2682
            ||                  || |||  ||||| || ||| |||||| ||||||||
Db     362 GGGACAGCATCGGCGGTCGACGGCTGCGGCGTCTTCGGCTACCCGGTTCAAGGACCCCG 421

Qy    2683 CCGCCTTCGCGTCCCGGCGAGGTCGACGCGTCGCGGCGTCTTCAGCCCGGGCCGTCC 2742
            ||                  ||||||| |||| |||| ||| || ||||||||
Db     422 AGCGGTACGGCGTCGCCGCCCAGTGGTGTGGGTGAAGCGGGCAGC 481

Qy    2743 CCGACGTCCGGCGTCGCTGCTCGCGCTCATCGGCGCTGTACGGCCGTCG 2802
            ||                  || ||| |||| |||||||||| ||||
Db     482 CCGTCAAGCGGCCGTCCAACTCGGCGTCACCGGCCGCTCTACCTCTACGACAACGACGTCG 541

Qy    2803 ACGAGGCGTACGGGCCGATCACCCCCTCCGCCGGCTGGAGAATCACCCACCGCGG 2862
            ||                  || || |||| |||||||||||||||||||||||
Db     542 TCGACATCGCCAAGAACATCCGGCCCTGCGGGCGCGCCGCGGAGCTGAGATCACCGACGTCA 601
```

```
Db    1269 TGCAATGACCAGTGGTGGCAGCAGGCGGATGACGTTGCCATCGGTGCCGCAAGTCAGCAG 1210
Qy     837 CCCCAGGTCGAGGCGGTGGTGTAGACGCGGCTTCGGGGGCTGCGGCGGGTGCCCGGCC    896
Db    1209 CAGCACCACCTTCAGCTTTTGCGCGCCGCAGCCACTGCGGGCGGGTTAAAGCTGCGTTCGGGGCG 1150
Qy     897 GACGGCGTCGGTGACGAACTCCAGGCCCCACAGTCCGAGGCGGCCGGTACCTGGCCGAC    956
Db    1149 GCCGGTAGCGTCGATAAGCTCAATGGCCATCATTCGCGCGGCACCGGCCGAGGATTTCGGCGAC 1090
Qy     957 CTCGGGGAAGCGGACTCCAGGCGGCGCAGCCGTCCTCGATGAGCTCGGCCGAGGACGCCG 1016
Db    1089 CTCCGGAAGGCACTCAGCTGCGCG--AATTCATCGCGGATGATGTCTCGATTTCTTG 1033
Qy    1017 CACGGCGGTCGATCAGCCGGTCGCGCTCGACGACCTCAGCGTGGCGCGCGCGCAT 1076
Db    1032 CGGCGGGTCTTAAGGTCGGCTTGTTGTTCCATCACTTCAATGGCTGCCGCCGGCGCA    973
Qy    1077 CCCCAGTGGTTGCTCGCGTACGTCGAGGCGTACGCCCCGGGGTGGCCGCCTCCGGCCTG 1136
Db    1136 AGCAACGGGGTTCCGCCGTAGGTTCCCGGACGCGCCGGG------GCCGGGTGCGTC    919
Qy    1137 CGCAGCTTCCGCGCGTCCGGCCACGCAGGGAATCCGCTCGGCGGTGCCCTTCGA 1196
Db     918 CATGATTTCTGCGGCGGCCGGTCACTGCGGATAGTGGCATGCCCCGGCGATGCCTTTTGC    859
Qy    1197 CAGCAGTGTGATGACGTCGGGGATCACCTTCGCTGGCAGGAAGGCGCCCGGTGCG 1256
Db     858 GGTGGTGATGACGTCGGTGAGGACCTCGTCGGCGACGAGCACGCCCGTCCCGGCCAGGCGCC    799
Qy    1257 CCCGCCGCCGGTGAGGACCTCGTCGTGCCGCTGGGGGGGATCATCGGGGGCGGTGCGGGGG 1316
Db     798 CAGGAAGCCAGATTGGATTTCATCGGCGGGGGGCCGATGAACACCACGTCGTTTCTCGCGGCACCAGGT 739
Qy    1317 GGGCGATCCCGCTCCCAGTAGCCGGCGCGCAGATGACGGCGTCGCCGCGGGCCGAGGACGCGG 1376
Db     738 GGAAATGGCTGCGAGGAATCCTGTGCGGGGACGATGAATCCGCCTTCGCCGTCGGATCGG    679
Qy    1377 TTCGAAGCCAGGGGCGCGAGACGTTGGGCTTCTCCGCGATGTGCCGGCCACGAGGGTCGC 1436
Db     678 TTCAATGACCACCGCAGGCGCAGGTTTTCGGGCTCCCAACCTGGATTCGATCACCGGAGATCGC    619
Qy    1437 GCACCGACACCTCGCACGACGAGGGGTAC---TCCAGGCCCAGGGGACCAGCGGTAGCCCAGTAGG 1493
Db     618 GCGCTCTGCGGCCTTCCGGCGCCGGACAGTCCGTCGCCGAGTGGTAAGACATTGTGCACG    559
Qy    1494 GGCTGTAGCCAGCACGCTGTTCGCCGCTGGTGCGGCCGGTCGTGCGGTCCGTGCCAGTGACCAG 1553
Db     558 GTAGACGTCTGCGCGGTCGCGCTAGTGTCCGGAATCCGGACTTGTATGGGCGGTTCTTCGGCGGTCAT    499
Qy    1554 CATCCGGCGCGGCCCATGGTCTTGCCGTGGAAGCCCGTGGCCGCAGCGGCGCAGAACCCGGTTGCG 1613
Db     498 CGCCATGGTGAGGTTGGTGGTCCGTCCGTGGTACGGCGT------TGTCGAACACCACGACCGC    445
Qy    1614 GCCCGGCGGGCGGGGCGGTCGCCTGGACGACCGCAGGGGGCGGCCTCGACCACCTCCCGCGCCGT 1673
Db     444 GCCCTTGCCGGTCGTGAGGCCGTGCGGCGGCCCCTTGACGCGTTTTCCACGGCCTTCGGCGGCCAGA 385
Qy    1674 GGAGAAGAAGGCGTAGGTGTCGAGCTGTTCGGGACAGCCTGCGGCGAGCAGTTCCAGCAG 1733
Db     384 GTTAAACAGCGCGCGCTCTCTGTCTGCGATCCGGCGGTCAAGGCGGTCTTGAGTCTCTCCGC    325
Qy    1734 GCCGGCGCGG-------TCCGGCGGTGCGCTTCGTCGTGCTGTCGCGGCCGCGGGGC 1784
Db     324 CATGCGCCACGTAAGGTTCATTAAGGTGAGACCATGGCACGCAGCAGGTGTGGGTGAAGCGGGCAGC    265
Qy    1785 CTGGGTGGTGTCGAGTGCCTCGACGACCTCCGGTCGGCGCCGACGACCTGCGTCGGGT 1844
Db     264 TGGCGGCCGGACGGCGTTCCGCAGGGTACTCGGTTGCCGTCCAGGTCCGTCAGAACGGGACCCGTCC 205
Qy    1845 CCCCGGGCCGCCGAAGTCGGAGGTCGATGAGAAAACGTTGCCGTCCGTCAGCCAAGATGCCACCGTCTGC 1904
Db     204 GCCGAGGCGCCAGGTCGATGAAAACGTTGCCGTCCGTCCGTCTCAGCCAAGATGCCACCGTCTGC    145
```

Qy  2863  TGCAGTGGATGATGACCGGGGGCTGCCGCTAGGGCCGA---GACCACCACCGGCCCT  2919
Qy  2863  TGCAGTGGATGATGACCGGGGGCCTGCCGCGTGTACGGCCTTCAGCCCGGCGCGTCC  2802
          ||||| |||| | ||| ||||  ||||||| |||||| | |  | |  |||  |||   |||
Db   602  ACCGCGTCTACCTCGGAGCGGGCGGGGCCGAACTCGTCAACCTGGCCGCGGCTTCGCCT  661
Db   482  CCGTCAAGCCCGGCGCTCCAACCTCGCCGCTCCAACCTCTACGACAACGACGTCG  541

Qy  2920  GGGCCGCAACCGGCAGCCGCGGGAGGACATGCTGGAGGTCAACCTGGACGGAC  2979
Qy  2803  ACGAGGCGGTACGGGCCGCCATCACCCCTCCGCCGGCGACTGGAGATCACCCACCCCG  2862
          ||| ||||||||||| |  | |  ||||||| | || ||  ||||| ||||||||||   ||||||||||
Db   662  GGCTGGACACCGGCACCCGCAGGGGTCGGA  721
Db   542  TCGACATCGCCAAGACATCGGCCCTGCGCGCGCGAGCTGGAGCTCACCGACGTCA  601

Qy  2980  TGGAGGGCCGCATCGAGGGGA  3000
Qy  2863  TGCAGTGGATGATGACCGGGGGCCTGCCGCGTACGGGCCGA---GACCACCACCGGCCCT  2919
          ||||| ||| | ||| |||| |   |  |  | ||  ||  |||
Db   722  AGGAGGCGGCAGGGCGTCTGGA  742
Db   602  ACCGCGTCTACCTCGGAGCGGGCGGGGCCGAACTCGTCAACCTGGGCCGCGGCTTCGCCT  661

RESULT 15

Qy  2920  GGGCGCGACACCGGCAGCGCGGAGGACATGCTGGAGGTCAACCGTCACGTCCTGGACGGAC  2979

US-09-988-384B-11
Db   662  GGCTGGACACCGGCACCCGCCACCCAGGACTCGCTCCTCCGGCCGCGCCCAGTACGTCCTGG  721
; Sequence 11, Application US/09988384B
; Publication No. US2003007382A1

Qy  2980  TGGAGGGCCCGCATCGAGGGGA  3000
; GENERAL INFORMATION:
;   APPLICANT: Sherman, D.H.
Db   722  AGGAGGCGGCAGGGCGTCTGGA  742
;   APPLICANT: Liu, H.
;   APPLICANT: Xue, Y.
;   APPLICANT: Zhao, L.

Search completed: May 29, 2003, 00:23:50
;   TITLE OF INVENTION: DNA encoding methymycin and pikromycin
Job time : 666 secs
;   FILE REFERENCE: 600.536US1
;   CURRENT APPLICATION NUMBER: US/09/988,384B
;   CURRENT FILING DATE: 2001-11-19
;   PRIOR APPLICATION NUMBER: PCT/US99/14398
;   PRIOR FILING DATE: 1999-06-25
;   PRIOR APPLICATION NUMBER: US 09/105,537
;   PRIOR FILING DATE: 1998-06-26
;   NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 11
;   LENGTH: 879
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-11

Query Match             2.3%;   Score 157; DB 9; Length 879;
Best Local Similarity  53.6%;   Pred. No. 2.9e-23;
Matches 397; Conservative    0; Mismatches 335; Indels    9; Gaps    3;

Qy  2269  TGAAGGCCCTGGTCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCGGTTCACCCACACCG  2328
          ||||||  | ||||||||| || ||||| | |  |  ||| ||| |||||   |||   |||  -
Db     2  TGAAGGGGATAGTCCTGGCCGGCGGGGAGCGGAACTCGGCTGCATCCGGCGACCTCGGTCA  61

Qy  2329  CGGCCAAGCAGCTGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTACGGCGTCGAGTCCC  2388
          | |||||| ||| |  ||  | |   |||| |||||| |    | |   |||| ||| | |||||
Db    62  TTTCGAAGCAGATTCTTCCGGTCTACACACAAACCCATGATCTACTACTATCCCGTCGGTTC  121

Qy  2389  TCGCCGCGGCGGGTGTCGGGAGGCCGGCGTCGTCGTGGGCGCGTACGGCGGCGGGAGATCC  2448
          | | ||  |   |   ||||||  | || | |  | |  |  |  |    |  -  |   | |||  -
Db   122  TCATGCTCGGCGGTATTCGCGAGATTCAAATCATCTCGACCCCCAGCACATCGAACTCT  181

Qy  2449  GCGA--ACTCACCCGGCGACGGCACCGCGTTCGGGTTACGCATCACCTACCTCCACCAGC  2505
          |  | |  | |||||   |||||||   ||  |  ||| |   | |||  ||| || ||||  -
Db   182  TCCAGTCGCTTCTCTGGAAACGGCCAGGCACCTGGGAATAGAAACTCGACTATGCGGTCCAGA  241

Qy  2506  CCCGCCCGCTCGGTCTCCGCGCCGCACCGGTGCCGCATCGCCCGCGGTTCCTGGCGACGACG  2565
          ||| | |||  |||| |||||||||||| |  | |   |  |  |  |  ||  ||||| ||||  -
Db   242  AAGAGCCCCGCAGGAATCGCGGAACGGCCACTTCGTCGCGGAGCCACCATCGGCCGACGACA  301

Qy  2566  ACTTCCTGCTGTACCTGGGGACAACTACCTGCCC---CAGGGGCGTCACCGACTTCGCCC  2622
          || | ||  || |||||  |||| | |||  || |||  |  |   |||    ||| ||||||  -
Db   302  CCTGCGCCCTGATCCTGGCGCACAACATCTTCACGGGCGCCGGCCTCTACACGCTCCTGC  361

Qy  2623  GCCAATCGCCGCCGCCGATCCGGCGGCCCGGCTCTGCCACCCCGGTCGCGGACCCGT  2682
          || |   | |   | | |||||||||||||| |||  |  || ||||| |||| ||||||
Db   362  GGGACAGCATCGCGGCCGCCTCGACGGCTCGCGCTCCTTCGGCTCACCCGGTCAAGGACCCCG  421

Qy  2683  CCGGCCTTCGGCGTCGGCGGAGGTCGACGCGGGCGACGGGAACGTTGCTCGCCGTTGGAGGAGAAAC  2742
          ||||| ||||||| ||||| ||||| ||||||||||||||||||| | |    ||||||| ||||||
Db   422  AGCGGTACGCGCGCTCGGCGTGGACGCGGGCGACGGGCTGACCGACCTCGTGCGAGAAGC  481

This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original
documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☒ FADED TEXT OR DRAWING

☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image
problems checked, please do not report these problems to
the IFW Image Problem Mailbox.

THIS PAGE BLANK (118 of 118)